ARI Research Note 96-15

Military Command Decisionmaking Expertise

James C. Deckert, Eileen B. Entin, Elliot E. Entin,
Jean MacMillan, and Daniel Serfaty
ALPHATECH, Inc.

Fort Leavenworth Research Unit
Stanley M. Halpin, Chief

January 1996

19960416 141

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

DTIC QUALITY INSPECTED 1

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

ALPHATECH, Inc.

Technical review by

James W. Lussier
Douglas K. Spiegel

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, January	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) August 1991-March 1994
4. TITLE AND SUBTITLE Military Command Decisionmaking Expertise		5a. CONTRACT OR GRANT NUMBER MDA903-91-C-0133 5b. PROGRAM ELEMENT NUMBER 0605502A
6. AUTHOR(S) James C. Deckert, Eileen B. Entin, Elliot E. Entin, Jean MacMillan, and Daniel Serfaty (ALPHATECH)		5c. PROJECT NUMBER M770 5d. TASK NUMBER 144 5e. WORK UNIT NUMBER C10
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ALPHATECH, Inc. 50 Mall Road Burlington, MA 01803-4562		8. PERFORMING ORGANIZATION REPORT NUMBER TR-631, ALPHATECH, Inc.
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-RK 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Research Note 96-15

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

COR: Rex Michel

14. ABSTRACT *(Maximum 200 words)*:

This report describes the development and validation of a theoretical framework for the investigation of tactical decisionmaking expertise. The theoretical framework was developed based upon interviews with U.S. Army command decisionmaking experts and a review of the literature on expertise. The primary means of validation was the conduct of a set of scenario-driven experiments using as subjects Army officers ranging in rank and experience from captain through General Officer. Three retired General Officers rated the level of expertise of 46 subjects independently based upon written products and videotapes. Nonmilitary researchers used the same set of products plus questionnaires to independently score a set of objective measures derived to test aspects of the theoretical framework. The three expert judges showed remarkable consistency in their independent ratings of the expertise level of the subjects. Many of the objective measures correlated with the experts' ratings. The objective measures did not, however, account for a significant enough portion of the variance to be, by themselves, reliable indicants of expertise. Suggestions for further research directions are presented in the conclusions.

15. SUBJECT TERMS

Expertise	Battle command	Decisionmaking
SBIR	Tactical decisionmaking	Command and control

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	111	

FOREWORD

A promising area of behavioral research in this time of reduced training budgets and diverse mission requirements is the investigation of battle command expertise. Discovering what qualities of knowledge, reasoning, and character distinguish those identified as experts offers a benchmark for selection and training. In addition to the benchmark, insights like these may guide the development of more efficient (i.e., better, faster, and less costly) training for battlefield command.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is involved in a major research thrust in the area of command leadership and decisionmaking. Related efforts include research into tacit knowledge, individual and shared mental models, situation assessment, decisionmaking and leadership styles, communication of commander's intent, and others. ARI is dedicated to helping the Army develop leaders for the 21st century.

ZITA M. SIMUTIS
Deputy Director
(Science and Technology)

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

This work was supported by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Fort Leavenworth Field Unit, Kansas. The authors would like to thank Mr. Rex Michel (ARI) for his energetic direction and hands-on participation and Dr. Stanley Halpin (ARI) for his continuous guidance and encouragement. We also acknowledge the mentorship of the retired Army General Officers who, while being remarkably patient with our own ignorance of real military operations, helped us develop our hypotheses and design the tactical scenarios, and served as "super-experts" in the project. The authors are especially grateful to the students and faculty from the School of Advanced Military Studies who made themselves available for the interviews, the workshop, the scenario development, and the Command Decisionmaking Expertise (CODE) experiments despite a hectic school schedule. Without their continuing participation, any theory of battle command decisionmaking will remain just that--a theory. Finally, we sincerely thank all the U.S. Army officers who volunteered to participate in this study and share their battle command expertise with the research team.

MILITARY COMMAND DECISIONMAKING EXPERTISE

EXECUTIVE SUMMARY

Research Requirement:

The nature of expertise in military command has become a topic of increasing interest over the past decade. An important aspect of this expertise is the ability to rapidly and effectively make decisions in dynamic battlefield situations where important information is not available. We need to better understand what superior battlefield decisionmaking performance entails and the relationships between it and the knowledge, decisionmaking styles, and background experiences that contribute to it. Better understanding these things will contribute to improved selection, training, and aiding of battlefield decisiomakers.

Procedure:

In Phase I of this Small Business Innovation Research (SBIR) we developed a theoretical framework for investigating Miltary Command Decisionmaking (MCD) expertise based upon interviews with military practioners and review of the expertise literature. In Phase II we sought to validate and enhance the theoretical framework through experimentation, field observation, and conduct of a joint military and researcher workshop. Throughout this project, three retired General Officers with recognized expertise as tactical decisionmakers acted as consultants and analyzed the experimental data.

Findings:

Our experimental procedure using tactical scenarios, controlled procedures, and expert raters proved successful in measuring MCD expertise. The expert raters were remarkably consistent in their independent ratings of the expertise level of individual participants. Many of the objective measures developed to evaluate aspects of our theoretical framework correlated with the expertise ratings of the expert raters. They did not, however, account for a sufficient portion of the variance to be, by themselves, reliable indicants of expertise, but some combination of the variables may yield a reliable prediction of expertise.

Utilization of Findings:

Probably the most immediately useful aspect of the findings is the comparison of the expert raters' rating justifications of low-expertise participants versus high-expertise participants. These comparisons are described on pages 50-53 of this document. The consistent themes that emerge from the comparison offer measures that might be applied to samples of decisionmaking in the classroom or even in field exercises. Those objective measures that correlate with the expert raters' comments provide an additional level of detail regarding how to measure these qualities. Overall, the findings suggest that command decisionmaking expertise <u>can</u> be reliably judged. What needs to be done is to validate more objective measures of its components and their interactions.

MILITARY COMMAND DECISIONMAKING EXPERTISE

CONTENTS

LIST OF TABLES

Page

LIST OF TABLES (Continued)

LIST OF FIGURES

MILITARY COMMAND DECISIONMAKING EXPERTISE

Introduction

Motivation

In the fields of decisionmaking, planning, and problem solving, expertise has long been one of the most difficult concepts to understand, capture, and quantify. The challenge is even greater for those seeking to understand expertise in military battle command, where decision tasks reflect the high levels of complexity, dynamism, and uncertainty inherent in tactical and operational environments. In the last few years, a great deal of interest has been generated in both the academic and military communities about the nature of expertise in command. An essential component of this expertise is the ability to make and implement decisions in a timely, efficient, and effective manner, most often with very limited information, in an increasingly fluid and multidimensional battlefield. We call this ability military command decisionmaking (MCD) expertise.[1]

We seek a theory of MCD expertise to provide a framework for analyzing and understanding the relationships between superior tactical and operational performance and the various factors affecting that performance. The theory must generate testable hypotheses to guide subsequent theoretical and empirical research. Once empirically validated, the theory will specify the components of expertise and have implications for how best to develop command expertise.

A clear framework describing the nature of expert command decisionmaking will have implications for developing training methods and materials, planning aids, and decision-support systems, as well as for testing, selecting, and evaluating personnel. Training practitioners and researchers in human performance seek to understand the cognitive processes underlying expert decisionmaking for the purpose of improving methods to educate and train nonexperts in more efficient and effective ways. On the other hand, the technology-oriented research community is looking at ways to quantify expert cognitive characteristics and decision rules and organize expert knowledge for building advanced data bases, planning aids, and expert systems that could provide support to decisionmakers in realistic environments.

A significant challenge for research on MCD expertise is the difficulty associated with assessing the degree of such expertise in an individual. The domain of the military commander is highly uncertain, multidimensional, affected by extraneous, uncontrollable factors (primarily weather and terrain), and dangerous. This is in contrast to many domains for which significant research in expertise has been conducted, such as chess. In chess the board ("ground truth") is always visible, the pieces have limited moves, and games can be played inexpensively and, if desired, repeated from any point. Thus, while in chess a system of national and international rankings of expertise (supported by significant numbers of scheduled matches and challenges) is available, no similar approach for rating MCD expertise is practical. Easily available metrics such as military rank and years of service, analogs of which have proven useful in rating expertise in fields such as medicine, have not proven to be reliable predictors of military command performance (the Little Big Horn is one of many historical examples). On the other hand the evidence for expertise associated with victory in an actual combat situation is debatable because of the many possible extraneous factors involved. Thus, a means for reliably assessing the MCD expertise of an individual is a necessary starting point for meaningful research in this area.

[1] We use the term command decisionmaking to indicate both operational (Corps and above) and tactical (division and below) applicability.

Approach

Our predecessor Phase I Small Business Innovation Research (SBIR) project was an attempt to shed some light on the decision and command strategies used by expert military decisionmakers in tactical situations. In Phase I we developed an initial theoretical framework and proposed a set of hypotheses pertaining to tactical expertise. (Hereafter we use the term tactical in the broad sense of relating to the employment of forces in combat.) The theoretical ideas and findings of the Phase I work were obtained through an extensive literature review and a set of interviews with commanders at the major and the general officer levels.

The Phase II effort reported here significantly expanded the work performed in Phase I, seeking to validate and enhance a practical theory of MCD expertise through a carefully designed research program of consultations with expert military commanders, direct observations and evaluations in exercises, and realistic and rigorous experiments. This approach combines the unique, invaluable perspectives and insights of experts acknowledged by the military community, the realism (hopefully fostering natural or near-natural behavior) offered by exercises, and the scientific control (to minimize extraneous factors and effects) and statistical analyses (arising from multiple replications and multiple subjects) allowed by carefully planned and designed experiments.

Key Activities

Theory Development

MCD expertise cannot be easily defined or explained. Phase I of this research effort drew together the cognitive science literature on expertise with a series of semi-structured interviews with military commanders in order to develop a theoretical framework for understanding tactical decisionmaking expertise (see Serfaty, MacMillan, and Deckert, 1991). This theoretical framework drove the design of the Phase II experiments and provides a structure for interpreting their results. We review our theoretical framework in the next section of this report.

Field Observations

In order to help us refine our theory of MCD expertise, to assess the feasibility of doing systematic observations in a military exercise, and to help us in preparing realistic experimental materials and procedures, we observed a Battle Command Training Program (BCTP) Battle Command Seminar and a BCTP Warfighter exercise. The observation activities and conclusions are discussed in the interim report for this project (Deckert, Entin, Entin, MacMillan, and Serfaty, 1992). Our observation of the Warfighter exercise provided support for our theoretical framework for MCD expertise, but we concluded that a warfighter exercise does not provide a suitable forum in which to systematically test hypotheses derived from our theory. The discussion of our field observations is included as Appendix A of this report.

Workshop

During Phase II of this project we designed, coordinated, and helped conduct a two-day workshop on Military Command Decisionmaking. This workshop was designed to explore multiple aspects of expertise in military command decisionmaking. The workshop brought together representatives from the armed forces, academic, and R&D applied communities to explore emerging ideas and training principles for the development and enhancement of command decisionmaking expertise. Selected experts presented papers in four areas addressed in the

workshop: theoretical issues, simulation and training, doctrine, and total career development. Working groups in each area met during the workshop and then presented their conclusions and recommendations to the group as a whole. Documentation of the goals, participants, and activities of the workshop are discussed in *Developing Command Decision-Making Expertise: Workshop Report* (Serfaty, Deckert, Entin, Entin, and MacMillan, 1993).

Experiments

We conducted two experiments during Phase II of this project. We refer to these as the COmmand Decisionmaking Expertise (CODE) I and CODE II experiments. The CODE I experiment was conducted during the first year of Phase II. A report of the CODE I procedure and results is presented in our interim report (Deckert et al., 1992). In this final report we focus on the combination of the CODE I and CODE II results. The third section of this report describes our methodology for both experiments, and the fourth section describes the results. In the last major section we present a summary of our findings and conclusions.

Recommendations

The last section of this report presents recommendations for future research on MCD expertise and recommendations for potential applications of our methodology for eliciting and assessing and training MCD expertise. Suggested areas for future research include additional experiments that build on the CODE experiments and additional analyses of the CODE data. Suggested application areas include training in the development of MCD skills, the assessment of MCD performance, and the assessment of the effectiveness of wargaming simulations for increasing MCD expertise.

Relevance to the Army

Despite the explosion of information, computer, and communication technologies, the increasing complexity, fluidity, lethality, and dimensionality of the modern battlefield make extreme demands on the decisionmaking skills of military commanders. Yet expert military commanders manage to maintain an accurate image of the tactical situation and make rapid and effective decisions under conditions of high stress and uncertainty. It is imperative that we understand how this decisionmaking expertise can be affected by the career-development process, and develop the most-effective means of building expertise. Reductions in military spending, and the associated reductions in manpower levels, make this situation all the more critical now and in the foreseeable future.

Report Organization

In the second major section we review the theoretical framework developed in Phase I, and enumerate verifiable hypotheses suggested by that theory. In sections three and four we discuss the CODE experiments conducted in Phase II. Section three presents the rationale, goals, hypotheses, method, measures, procedure, and data-reduction process for the experiments. Section four discusses the experiment results. The final section presents a summary of the report, conclusions, and recommendations for applying the results.

There are two appendices to this report. Appendix A contain a report on our field observations at a training program and a command-post exercise. Appendix B contains copies of the experiment materials used in the CODE experiments.

Theoretical Framework

What is the nature of MCD expertise? Phase I of this research effort drew together the cognitive science literature on expertise and a series of semi-structured interviews with military commanders in order to develop a theoretical framework for understanding tactical decisionmaking expertise (Serfaty, MacMillan, and Deckert, 1991). This theoretical framework, which we review briefly in this section, drove the design of the CODE experiments and provides a structure for interpreting their results.

Premise: Mental Models

The underlying premise of the theoretical framework is the cognitive science concept of "mental models." Mental models are our internal representation of the external world. We suggest that an expert commander has a mental model of the tactical situation that differs in measurable ways from that of a less expert commander.

The Phase I report reviewed literature on expertise and the use of mental models to represent expert knowledge in a variety of fields from chess to medical diagnosis. We also reviewed literature on military command decisionmaking and on tactical expertise. This expertise literature suggests that the expert's initial understanding of the problem or assessment of the situation is a critical part of decisionmaking expertise. Experts begin their problem solving by assessing the situation rather than plunging immediately into detail, and a large component of expertise is in knowing how to frame the problem.

The literature on mental models suggests that such models can provide a mechanism for representing the expert's understanding of the situation. The expert's memory consists of an extensive array of "patterns," with information items grouped together and indexed by their relevance for problem solving in the domain of expertise. We suggest that the expert's pattern-indexed memory supports the construction of a better initial mental model of the situation. The expert can retrieve a problem representation structure from memory that is similar to the problem at hand in a way that facilitates problem solution. We concluded, based on the literature, that the expert's mental model of the situation is a key factor in MCD expertise.

The literature does not tell us what, exactly, is contained in the MCD expert's mental model of the situation that makes it superior. It also does not provide insight into how the military commander uses his[2] mental model to deal with the uncertainty of a tactical situation. To gain insight into these issues, we conducted a series of interviews with military commanders.

Interviews With Military Commanders

We conducted interviews with military commanders (three retired general officers and three majors) at (presumably) several levels of expertise. These interviews used a semi-structured format in which the interviewees were presented with a scenario featuring an unexpected "critical incident" and asked for their reactions to the incident. The commanders were also interviewed at length concerning their views on the nature of MCD expertise. These interviews provided us with more specific material on the nature of decisionmaking expertise in a military context, allowing us to identify its similarities to and differences from expertise in other fields.

---

[2] Unless otherwise stated, whenever the masculine gender is used both women and men are included.

The interviews produced the following observations, discussed in more detail in Serfaty et al. (1991) and in Serfaty and Michel (1990):

1. Experts have a flexible plan. One of the key differences that we observed between the responses of less expert and more expert commanders to the critical incident was their attitude toward the plan. Nonexperts tended to be locked into the plan, while experts looked at the plan as a foundation on which to build contingencies. The expert commanders were also more aware of the boundaries of the plan and the impact of possible plan modifications on adjacent units and subordinate headquarters.

2. Experts learn from their mistakes. Experts showed the ability to learn from past decisions and to make appropriate changes in future decision strategies. In contrast, nonexperts appeared more interested in rationalizing or defending past decisions than in learning from them.

3. Experts never forget the enemy. Experts allocated a large amount of attention and mental energy to the enemy. A key portion of their planning process was devoted to guessing his intent, "reading" intentions from his actions, and maximizing the damage they could inflict on him.

4. Experts seek (disconfirming) information. Experts preferred to gather information in a proactive way. Most experts emphasized the importance of information pull as opposed to information push. Expert military commanders know that, in a hostile environment, things rarely go according to plan. Their awareness of an intelligent enemy apparently induces them to look for evidence of deceptive operations, and to prepare for these contingencies ("a good warrior is also a worrier").

5. Experts include the human element in their plans. The expert commanders focused a major part of their efforts on understanding the human element in a tactical situation. Experienced field commanders seem to have a clear mental picture of human performance in combat, the effect of increasing casualties, and the impact of fatigue on troop morale and effectiveness.

6. Experts build teams. Expert commanders reported that they were quite aware of the team around them and planned to make effective use of its capabilities. The concept of teamwork seemed to be essential to their command philosophy.

7. Experts act more effectively and faster under uncertainty. A key factor that differentiates experts from nonexperts is the way in which they deal with uncertainty. First, experts have a more sophisticated understanding of uncertainty. They look at it as a dynamic process, evolving with time and emanating from different sources (e.g., nature, terrain, enemy). They carefully select the tools used to reduce that uncertainty, being acutely aware that doing so can also give information to the enemy as to one's own plans. They also know that the quest for absolute certainty is doomed to failure and carries a high cost in timeliness and speed of their decisions. They have, therefore, a higher tolerance for exogenous uncertainty and can manage it within acceptable levels of stress. Figure 1 is a graphical representation of our observation from the interviews that the experts were willing to make decisions more quickly, based on more uncertain information, than the nonexperts. Figure 1 also captures our observation that experts achieve a faster rate of reduction of uncertainty because they are able to recognize critical elements of information and seek them in a proactive fashion ("directed telescope").

8. Experts explore an option in depth. A central observation from the theory of recognition-primed decisionmaking (RPD)(Klein, 1988) is that expert decisionmakers "recognize" a situation (from a past experience), and this recognition generates a course of action. On the other hand, theories of normative decisionmaking prescribe the systematic generation of multiple options and the selection of the best option according to some criterion of perceived

effectiveness. Army doctrine recommends the latter, but observation in the field seems to indicate that the former is the prevalent behavior in tactical environments.

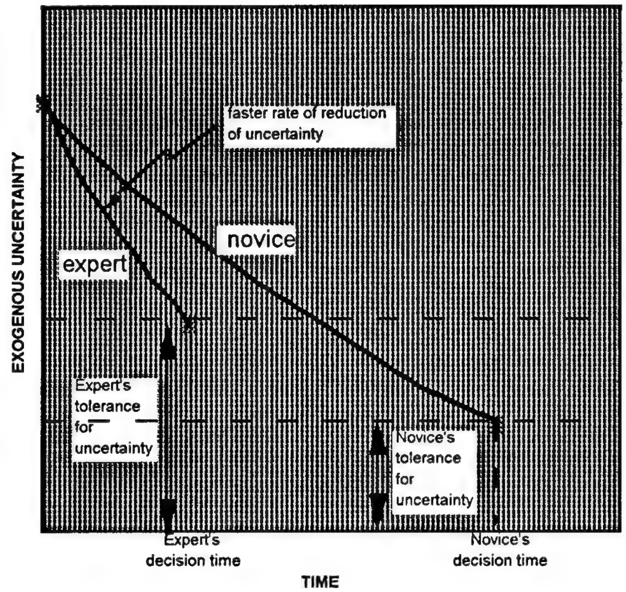


Figure 1. The decision to decide.

Based on our interviews, both patterns of behavior exist in realistic settings and are sometimes in use simultaneously. The prevalence of one over the other depends strongly on the nature of the mission, the task at hand, the time available to make a decision, the stage of the process (on-line decisionmaking vs. off-line planning), and the expertise of the decisionmaker.

Most activity that we observed for expert commanders involved option exploration rather than option generation and selection. Thorough analysis ("what if" questions) was considered essential for the adoption of an adequate option with multiple explored branches.

9. Experts organize and communicate their knowledge through "war stories". War stories are a common feature of any discussion with an expert commander. During our interviews, commanders constantly used war stories and other analogies to illustrate ideas and express concepts.

The use of analogies and war stories during the interviews indicates an inductive reasoning process in the expert commander's problem solving. However, as discussed below, we observed that this intuitive process is only part of the expert commander's toolbox and that induction is used in a directed fashion.

10. Experts use both intuition and analysis in planning. The interviews indicated that expert decisionmakers are quite resourceful in their ability to switch between intuitive, inductive reasoning and analytical, deductive decision strategies. We observed expert military tacticians

using imagery and analogy (induction) to assess the situation and to recognize known patterns that matched their experience. Depending on the closeness of the fit, they then tried to complete the picture with specific information requests (directed telescope). Once a situation was recognized or at least categorized into a class of similar situations, experts used their analytical skills to break the problem into smaller chunks of information and explore in a systematic and detailed way (deduction) the consequences of the hypothetical decisions and actions that would follow. After an option had been chosen and its branches explored, experts switched back to their inductive reasoning mode to picture the new situation as a whole (augmented by a mental simulation of the impact of their course of action) and matched it against the mission requirements. This three-stage process could be iterated several times or speeded up depending on the time constraints.

The experts interviewed seemed to be acutely aware of the limitations and domains of application of their mental models. They described a process by which expert commanders constantly test and refine their models against hypothetical situations; they mentally "see" the situation and "play" with various missions. In addition, expert commanders look at a tactical situation and measure it first against high-level concepts such as fluidity of forces, observability of operations, accessibility of terrain, etc., before going into specifics, such as METT-T.

We call this transition between inductive and deductive reasoning the "hourglass" model of the planning process, as illustrated in Figure 2. It is an extension of the RPD model proposed by Klein (1988).

Framework Based On Literature and Interviews

Our next step was to integrate our literature review with the interview results to develop a framework for understanding MCD expertise. In some cases the theory and research from the literature on expertise and mental models fitted well with the behavior observed and the comments made in the interviews. Other aspects of expert behavior observed in the interviews did not fit so well with the previous literature, and required the development of extensions to existing theories of mental models and expertise.

Our framework for understanding MCD expertise builds on and expands the hourglass model to describe the process by which mental models are developed and used by the expert tactician. Figure 3 shows this expanded version of the hourglass model and summarizes evidence from the literature and the interviews that supports the proposed framework.

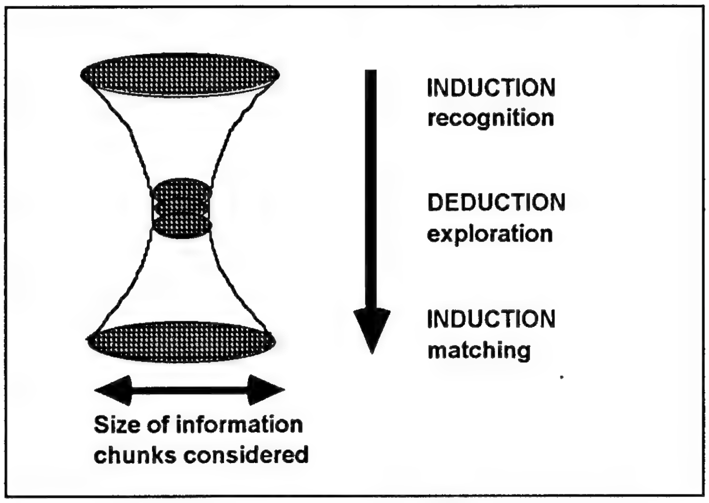


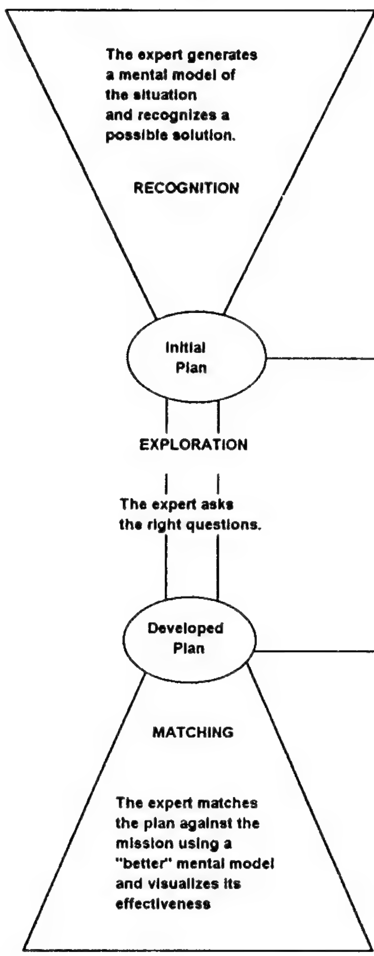
Figure 2. The "hourglass" planning process.

Recognition.

Based on the expertise literature, we suggest that expert tacticians organize their knowledge base in order to store a large amount of information in their domain, and they are able to "chunk" information by grouping details together into patterns. They store and retrieve information about their domain in a different way than nonexperts, and are able to very quickly bring to mind the relevant experience and information. This is consistent with Klein's RPD model and with our observation from the interviews that expert commanders store their knowledge in the form of war stories.

Cognitive science research indicates that human beings encode a variety of information about an experience, and can access their memory of it along many different dimensions. This memory retrieval does not appear to be a conscious process. Johnson-Laird (1983) argues that parallel processing occurs in memory retrieval, precluding the possibility of conscious awareness. Thus the expert commander may not be aware of how he accesses his collection of war stories, just that certain circumstances bring certain previous experiences to mind. This is not to say that experts are not able to explain the relevance of an experience once they have retrieved it, just that they aren't aware of the mechanisms of the retrieval process.

How does the expert's retrieval of information from memory support the construction and use of mental models? We suggest that the expert commander builds a mental model of the current situation and his plan for dealing with it. The first step is to retrieve relevant experiences, which are very specific, from memory. None of them will match the current situation exactly. Taken together, however, they give the decisionmaker a "schema" or skeleton that indicates which aspects of the current situation are most important. The decisionmaker can them proceed to fill in the empty "slots" of the schema through information gathering, analysis, etc.



	SUPPORT FROM LITERATURE	SUPPORT FROM INTERVIEWS
Experts draw on long-term memory - "war stories" Retrieve most appropriate "schema" for mental model to represent current situation Initial plan is suggested by the situation (recognized)	Experts have more information in memory, and organize it better than novices Experts have a pattern-indexed long-term memory Experts excel at initial problem understanding and situation assessment Experts use mental models to represent the situation Experts "recognize" possible solutions	Experts learn from past experiences Experts focus on enemy intent Experts store information in the form of war stories
The schema helps experts to ask the right questions during analysis Questions are driven by structure of mental model ("what is important for me to know?") The expert's mental model supports decisions under uncertainty	Experts use step-by-step reasoning in nonroutime tasks Experts use mental models to explore possible solutions in unfamiliar situations	Experts explore options once they are generated Experts seek disconfirming information Experts have a higher tolerance for uncertainty Nonexperts put more emphasis on operations and logistics to the exclusion of enemy intent and reaction
The expert visualizes how plan will play out on the battlefield. Will plan meet mission requirements?		Experts can deal better with unexpected changes in their plan Experts have a better sense of the boundaries of their plan Experts have a good grasp of the state (and impact) of the human element Experts maintain a common representation of the situation for their team

Figure 3. Framework for understanding MCD expertise

Exploration.

After the expert commander orients himself through retrieving relevant instances from his previous experience, he needs to explore the details of the current situation and the plan being considered. This is a conscious analytical process, and may be supported by computers and other planning tools.

What is the role of the mental model at this stage of the process? We believe that the decisionmaker's model, based on his experience, suggests what kinds of important information are missing and must be acquired. The model keeps the decisionmaker from being overwhelmed by details and "lost in the trees." Information gathering and an analytic, deductive effort fill in the critical empty slots in the schema.

How is MCD expertise manifested at this stage? The expert may have a better set of analysis techniques at his disposal. Also, because he has a better schema for the overall situation and the important missing links, he may ask better questions, i.e., make better use of the information-gathering resources and analytic techniques that are available.

Matching.

Once the decisionmaker has developed a mental model of the situation and filled in its holes through analysis and information gathering (to the extent possible), he can play out his plan of action and check for feasibility and mission effectiveness.

How is MCD expertise manifested at this stage? We hypothesize that the expert has a "better" model of the tactical situation and of the plan and therefore that he can do a better job of visualizing outcomes, problems, etc. The expert's model may be better along a number of different dimensions. Some hypotheses about the important dimensions for MCD are discussed below.

Theoretical Hypotheses About MCD Expertise

This subsection discusses the hypotheses derived for each stage of the framework, presenting the supporting evidence for each and indicating which hypotheses are not yet well supported by evidence.

Experts have a different memory structure than nonexperts. The expert maintains an extensive store of specific experiences; relevant experiences can be retrieved quickly.

These hypotheses are supported by the literature on expertise in a number of different areas. The literature suggests that experts store and retrieve information on the basis of its usefulness in their domain. The second hypothesis is also supported by our observation that MCD experts store and communicate information in war stories.

The expert's memory structure is used to generate a schema for the new situation.

We suggest that the expert retrieves from memory not just specific facts but a structure or schema for representing the situation that helps him to organize information and to understand its implications. This hypothesis is partially supported by the literature on expertise and mental models. The literature suggests that experts in a task domain excel at initial problem understanding, and that experts represent their understanding of the situation in a mental model. What has not been well explored is how this mental representation of the situation is tied to or generated from the expert's extensive experience base. Every situation is different, yet the MCD expert is able to retrieve a structure from memory that allows him to quickly assess the new situation.

The expert's initial schema for the situation includes a possible plan of action.

The idea that situation assessment and option generation are inextricably linked is supported both by Klein's (1988) observational studies of real-world command decisionmaking and by Kintsch's (1988) cognitive science model that treats understanding and planning as one process. The way in which the mental model of the situation is used to generate an option is not so well understood, however.

We suggest that a possible plan of action is inherent in the structure of the schema used to represent the situation, and that both are retrieved from memory as one "package." We hypothesize that the expert has a store of experience in memory that is indexed and accessed by the kinds of action that were effective in each kind of situation. That is, when the expert retrieves relevant experience from memory, it is of the form: "this is the kind of situation where xx plan of action is possibly a good option."

The expert's initial schema for the situation and the plan helps him ask the "right" questions and do the "right" analysis.

We hypothesize that the expert's mental model of the situation and the plan focuses him on the most important gaps in his information and suggests the areas where information seeking and analysis are most important. This hypothesis is based on our observation during the interviews that experts sought information in a different way than nonexperts. In particular, experts sought information that might disconfirm their understanding of the situation and the enemy's intent. We observed that nonexperts put more emphasis on exploring the operational and logistics details of a plan rather than considering the uncertainties associated with enemy intent and possible enemy action.

Experts build and use a "richer" mental model of the situation and the plan than nonexperts.

We hypothesize that the mental model of the situation and the plan that is built and used by the MCD expert is superior to that of the nonexpert. This hypothesis is based almost completely on the interview results; the literature on mental models and expertise has not dealt in any detail with the differences between the mental models of experts and nonexperts. From the interviews, it seems that the mental models of experts may contain facts, operating procedures, inter-system interactions (sometimes very complex and multidimensional), goals, constraints, and mission. Often these elements are concatenated into a complex chunk and stored in the form of a war story.

We believe that the mental models of experts are richer than those of nonexperts. This hypothesis takes two forms. First, the mental model of the expert contains more and different information than that of the nonexpert. We observed during the interviews that experts included detailed information about the enemy and the enemy's intent in their planning process, while

nonexperts did not. We also observed that the expert tacticians considered the physical and psychological states of the troops under their command during planning, and sought information about that condition, while the nonexperts did not. These are both instances in which the experts' mental models of the situation and the plan contained different information (and more information) than those of nonexperts.

Second, we believe that the mental model of the expert is more highly connected than that of the nonexpert. The elements of the expert's mental model of the situation and the plan contain connections that are missing from the nonexpert's model. For example, we observed in the interviews that experts had a better sense than nonexperts of the boundaries of their plans, and of the ways that changes in the plan would affect adjacent units and subordinate headquarters. The experts were also more aware of the connections among their own staff, and were more concerned than nonexperts with building and maintaining a common understanding of the situation and the plan among team members. Kahan, Worley, and Stasz (1989) discuss the critical need for the commander to communicate his "dynamic image of the battlefield" to his staff. This image, as described by Kahan et al., includes the contextual surroundings of the battlefield as well as military, political, and psychological considerations.

<u>The expert's mental model of the plan and the situation allows him to visualize outcomes.</u>

The use of mental models to visualize outcomes is a common theme of the mental models literature. We suggest that the expert commander uses a mental model to play out the planned actions and visualize their effectiveness in accomplishing the mission. This is consistent with comments made during the interviews that the expert must be able to visualize the battlefield, visualize the effects of actions that take place outside his unit's boundaries, visualize his position from the enemy's point of view, and visualize the situation in terms of action/reaction/action.

<u>The expert's mental model allows him to deal more effectively with uncertainty.</u>

The presence of uncertainty is a critical part of the MCD environment, and it was a recurrent theme of the interviews that expert commanders dealt more effectively with uncertainty than nonexperts. Experts had a more sophisticated understanding of the dynamics of uncertainty than nonexperts and were willing to make more-rapid decisions under uncertain conditions.

This is an area in which neither the mental models literature nor the expertise literature offers any useful guidance. We have developed the following hypotheses about how experts deal with uncertainty based on the behavior observed during the interviews.

First, we believe that the expert may perceive a lower level of uncertainty than the nonexpert because he has a better, more predictive mental model of the situation. For example, if the expert feels that he has a good model of the enemy's capabilities and the enemy's intent, then he may experience less uncertainty than the nonexpert because he feels confident that he has bounded the possibilities for enemy action.

Second, the expert may act to reduce critical uncertainties based on his mental model of the situation and the availability of information. For example, he may seek information that would disconfirm a key aspect of his model of the enemy's capabilities or intent. Also, he may have a better model of the dynamic nature of uncertainty, and may have a plan for reducing uncertainty in the future, e.g., he may consider at what point in the future more certain information will be available and do his planning accordingly.

Finally, the expert may have a mental model of the situation that includes uncertainty in a way that the nonexpert's does not. One interviewee commented that "the expert's parameters are wider." This hypothesis suggests that the expert is able to represent a range of possibilities in his

mental model of the situation. The literature on mental models usually considers them to represent specific situations rather than a range of possibilities. However, the expert may be able to chunk a range of possibilities together and deal with them as a single piece of information rather than as a number of separate possibilities. For example, his mental model may specify that "the enemy will probably come from the north along one of three routes" rather than "the enemy may come southeast on route X," "the enemy may come southwest on route Y," and "the enemy may come south on route Z."

<u>This broader mental model of the uncertainties of the situation supports the expert in developing a more robust and flexible plan</u>.

In the example above, if the expert's model represents the enemy as coming from the north along any of several different routes, then his plan should be robust enough to be successful under all of the possible avenues of approach. This hypothesis is consistent with the hedging behavior observed in experiments on multiple-option planning (Entin, Needalman, Mikaelian, and Tenney, 1988). The hypothesis is also supported by the observation from the interviews that experts have a more flexible attitude toward their plans and are better able to assess the implications of new information and unexpected events.

Summary

Figure 4 summarizes the major components of MCD expertise suggested by the theoretical framework. We believe that the MCD expert maintains an extensive store of specific experiences in memory (consciously recalled as war stories), supported by high-level principles, that allow him to very quickly develop a schema (an initial, incomplete mental model) for a new situation. This schema is associated with possible courses of action. The schema helps the expert to focus immediately on the most critical aspects of the situation, to ask the right questions, and to gather the most relevant information. The expert uses the information that has been gathered to build a richer mental model of the situation than the nonexpert. This mental model captures the dynamics of the situation in both space and time, and the expert uses it to visualize the outcomes of possible courses of action. We believe that the expert's mental model also supports more rapid decisionmaking under uncertain conditions, although the mechanism by which it accomplishes this has not been specified. The use of this richer mental model allows the expert commander to develop a course of action that is both flexible and robust. The COA's flexibility is the result of contingency planning and a careful exploration of the options and the associated "branches." A richer mental model allows the expert to simulate mentally the different ways his decisions may affect the situation ("what if" questions) and to prepare for deviations from the main COA. The COA's robustness reflects its resilience against uncertainty. A COA developed by an expert is robust enough to anticipate a wide range of situational variations. The combined qualities of flexibility and robustness result in the expert's taking effective action in the situation.

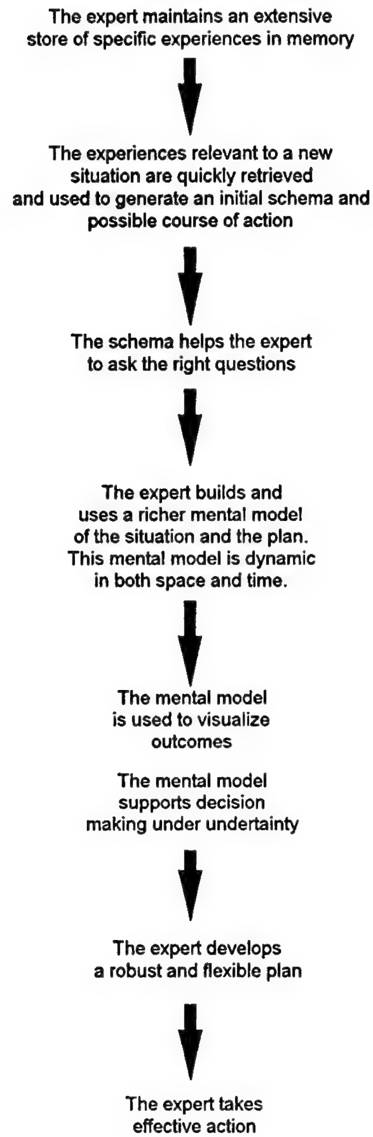


Figure 4. Theoretical components of MCD expertise.

Experimental Design

We designed the CODE I and II experiments to gather data on MCD expertise in a controlled laboratory-style setting in order to assess the adequacy and validity of the theoretical framework described in the preceding section. This section discusses the rationale for the experiments, the hypotheses tested, the design of the experiments, and the methods used to conduct the experiments and analyze the experimental data.

Rationale and Goals of the Experiments

The study of expertise faces many methodological challenges. Expert behavior in solving well defined, "toy" problems is easy to study in controlled experiments, but may bear no relationship to the competence of real-world experts. On the other hand, real-world expertise can be difficult to elicit, measure, and study under laboratory conditions. The complexities and subtleties that distinguish the true expert from the novice or nonexpert in solving real problems may be simplified out of existence in the lab due to limited resources, limited time, and a limited understanding of the subject area by the experiment designers. Field studies, in contrast, allow observation of experts solving real problems in natural environments, but are impossible to control in the experimental sense and allow no repetition of conditions, no systematic comparisons of expert and nonexpert behavior, and no statistical analysis of results.

The methodological challenges of studying expertise are even greater for MCD expertise. The true expertise of the commander is seen on the battlefield where no systematic observation or measurement is possible. Also, studying MCD expertise poses difficulties in the identification of experts. In areas of expertise such as medicine, individuals specializing in a particular field pass through a well-defined series of steps (e.g., medical student, intern, resident, specialized practitioner) and these levels, to a great extent, can be used to define their expertise. Because there is no "single career path" for the expert commander, there is no obvious way to identify expert commanders based on age, grade, etc. While it seems likely that specific types of training and/or experience are related to MCD expertise level, there is no consensus at the present time on which factors are most important.

The first major goal of the CODE experiments was to demonstrate that it is possible to define and measure MCD expertise in a reliable manner based on behavior that occurs under controlled conditions, i.e., conditions that can be repeated over multiple subjects. Our theoretical framework suggests certain behaviors we would expect to see on the part of the expert commander. We did not want to use these behaviors to define expertise, however, because this produces a circular argument in which an expert is defined as someone who exhibits those behaviors that are predicted by our theory of expertise. Instead, we asked three "super-expert" judges with extensive tactical experience (their qualifications are discussed later) to provide independent ratings of the expertise of each subject in the experiment. We did not define the meaning of expertise for these judges, but asked them to base their ratings on whatever criteria they individually felt to be indicative of MCD expertise. The extent to which these judges were able to agree on a rating for an individual's expertise level based on information from the experiment, and the extent to which these ratings are stable for an individual but vary over the subject pool, provide evidence on: 1) whether there is a quality called "MCD expertise" associated with an individual, and 2) whether that quality can be elicited in a laboratory experiment.

The second major goal of the CODE experiments was to determine whether the observable behavior of expert commanders is consistent with the description of expertise in our theoretical framework. This requires an assessment of the extent to which the observable behaviors and measures that we would expect to be associated with expertise level according to the framework are in fact correlated with the expertise-level ratings of the judges. Do experts (as identified by

the judges) behave in the way that we would expect based on our theory? We developed a large number of secondary expertise measures (i.e., measures that we expected to correlate with expertise level) based on the framework, and tested their correlation with the judges' ratings. These secondary measures are based both on the observable behavior of the subjects as rated by ALPHATECH staff and on the subjective reports (both written and verbal) of the experiment subjects.

The CODE experiments had a third, more descriptive purpose as well. We asked our expert judges to provide written explanations of the basis for each of their expertise ratings. This narrative material provides insight into the judges' concepts of MCD expertise. It also is useful in assessing the adequacy of our theoretical framework and in identifying ways in which the framework can be expanded or improved.

Hypotheses

The hypotheses tested in the CODE experiments fall into two major groups: methodological hypotheses about the ability to measure MCD expertise, and theoretical hypotheses about the nature of that expertise.

The methodological hypotheses are fundamental to the concept and design of the experiments. We hypothesize that there is a quality called MCD expertise, that this quality is associated with individuals, that this expertise can be elicited using written materials and maps to pose a tactical problem in a laboratory setting, and that this expertise can be reliably measured by well-qualified judges. These fundamental methodological hypotheses form the foundation for the CODE experiments, and must be supported by the data before we can take the next step of testing our theoretical hypotheses on the nature of MCD expertise.

The theoretical hypotheses of the experiments concern the relationship of a variety of secondary measures to expertise level, as suggested by the theoretical framework. As discussed earlier (see Figure 4), this framework suggests that an expert commander has an extensive store of relevant experience in memory and is able to draw on this store of experience very quickly to generate a schema for a new situation. This schema helps the expert ask the right questions and gather the right information. Based on this information and on information from memory, the expert builds a richer mental model of the situation than the nonexpert, and uses this model to visualize the outcomes of alternative plans. The mental model also supports decisionmaking under uncertainty. Using the mental model, the expert develops a plan that is more flexible and more robust to possible changes in the situation than the plan developed by a nonexpert.

In designing the CODE experiments, we wanted to create situations in which the expert commander would have a chance to exhibit all of the behaviors described in the previous paragraph. We wanted to be able to observe a subject's initial reaction to a tactical situation, the questions asked in the effort to gather more information, the process of developing a COA, the COA that was developed, and the reaction of the subject to new, unexpected information that might cause a change in that COA. The experimental design described below provides, for each tactical situation:

- an initial description of the situation,

- an opportunity for the subject to volunteer an initial COA or provide an initial COA at the prompting of the experimenter,

- the opportunity to ask questions,

• time to prepare a final COA after the questions have been answered,

• the arrival of new information relevant to the scenario [CODE I only], and

• an opportunity to react to that new information by changing or modifying the COA if necessary [CODE I only].

We were able to observe the subjects' behavior throughout this process and to administer questionnaires at several break points.

Based on the theoretical framework, we defined a series of expected behaviors that could be observed during COA development, and a series of questions to elicit the subjects' perceptions at each stage. These observable factors and subjects' perceptions produced a set of secondary measures of expertise, as shown in Figure 5. We hypothesized that each of these measures would be correlated with MCD expertise as rated by the judges.

The first theoretical component of expertise is the retrieval of specific experiences from memory, and the use of these experiences to generate an initial schema with an associated plan of action. While we could not measure the contents of subjects' memories directly, we did ask them if the situation as described reminded them of any previous experiences. We also expected to see the more-expert subjects generating an initial COA more quickly than the less-expert subjects, and we expected that the initial COAs generated by more-expert subjects would be more detailed and contain more contingencies than those generated by the less-expert subjects. If the initial COA is more "on target" for the more-expert subjects, we expected to see more agreement between the initial COA and the final COA (prepared after questions were asked) for the more-expert subjects.

We expected to see a difference in both the number and the relevance of the questions asked by the more-expert and less-expert subjects. The theory suggests that more-expert subjects use their initial schema and associated COA for the situation to quickly identify important gaps in their knowledge so that they can obtain the most-critical information. We also expected that, if the more-expert subjects were guided in their questions by the information that was missing from their initial schema, we should see greater use of the responses to their questions in the final COA (i.e., the expert goes immediately for the critical information without wasting time on details that are not relevant to the plan).

Although we could not observe the experts' mental model directly, the hypothesis that the expert builds and uses a richer and more dynamic mental model of the situation generates a number of expected secondary measures. First, we expected that the more-expert subjects would "visualize" the situation in a very concrete way. Thus, we expected them to rely more heavily on the map, as a visualization aid, while studying the situation and developing their mental models. We also expected the COAs prepared by the more-expert subjects to take account of complex timing issues to a greater extent than those prepared by the less-expert subjects. We expected

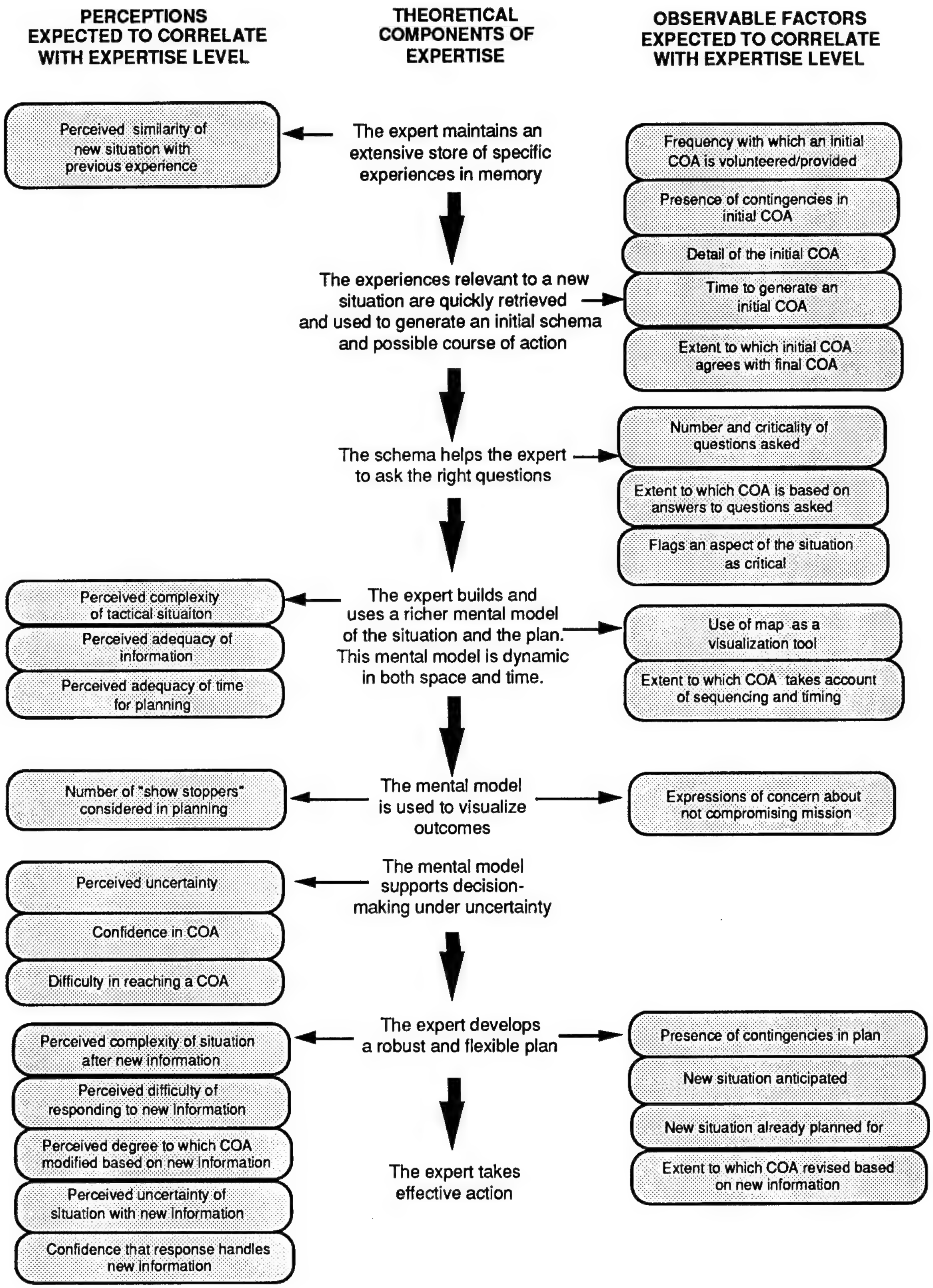


Figure 5. Secondary measures of expertise based on theoretical framework.

would differ from the less-expert subjects in their perception of the complexity of the situation, and in their evaluation of the adequacy of the time and information available for COA development. We were unsure, however, of the direction of these relationships. Does the expert perceive the situation as less complex because of his better mental model, or as more complex because he is aware of issues (such as synchronization and proper employment of the battlefield operating systems) that are not apparent to the nonexpert?

The theory suggests that the expert uses his richer mental model of the situation to visualize outcomes while developing a plan. We have two measures related to this hypothesis: the extent to which the subjects vocalized concerns about the "end state" of their plans as it related to the mission, and the number of impediments or "show stoppers" that they reported considering during the planning process. We expected that the more-expert subjects, because they had a greater ability to visualize outcomes, would be more likely to mention the importance of not compromising the mission, and would report having considered more potential impediments during planning than the less-expert subjects.

We believe that the expert is able to make decisions more effectively than the nonexpert under uncertainty, but we had multiple hypotheses about the way in which the expert's mental model supports this ability: the expert perceives less initial uncertainty; the expert acts more effectively to reduce uncertainty; and the expert's mental model allows him to act under uncertainty, i.e., to take an action that will be effective across a range of possible conditions. In order to gain insight into the way in which the more-expert tactician deals with uncertainty, we asked subjects about their perceived uncertainty after they were presented with the tactical situation, their uncertainty after developing a COA, their confidence in the COA they had developed, and their difficulty in developing the COA.

Based on the theory, we expected that the COAs produced by the more-expert subjects would contain more robust and flexible plans than those produced by the less-expert subjects. We expected to be able to detect this in a number of ways. The first was the presence of contingencies in the COA: we expected the more-expert subjects' COAs to have more contingencies. We also expected to see differences in the way that the more-expert and less-expert subjects reacted to the new information presented to them after the COA had been developed. We expected the more-expert subjects to have anticipated and already planned for the new situation, and to make fewer changes in their COAs based on the new information than the less-expert subjects. We also expected that the more-expert subjects' perceptions of the situation after the new information was introduced might differ from those of the less-expert subjects, including the perceived complexity and uncertainty of the new situation, the difficulty of responding, the perceived need to modify the COA, and the confidence that the modified COA can handle the new situation. Again, we were uncertain of the predicted direction for these differences in perception: are experts more certain and more confident because they have developed a more robust COA, or are they less certain and confident because they have a clearer picture of the situation and have considered more of the possible impediments to the plan?

The remainder of this section explains how we tested these ideas about MCD expertise in the CODE experiments. It describes the design and conduct of the experiment, the method used for the judges' ratings, the operational definitions of the secondary measures, and the methods used for data collection and data analysis.

Method

The CODE I and CODE II Experiments

The CODE I and CODE II experiments were very similar in purpose and design. CODE I was designed to assess our proposed method for evaluating MCD expertise and to test the hypotheses developed from our theoretical framework. CODE II was a follow-on to the CODE I experiment, intended to increase the number of higher-expertise subjects. The distribution of expertise levels in the CODE I experiment, in which only one subject was above the rank of major, was negatively skewed. We hoped that by including a number of higher-ranking subjects in the CODE II experiment we would obtain a larger sample of subjects with higher levels of expertise. With a more symmetric distribution of expertise scores, we felt we could conduct more sensitive tests of our hypotheses.

Because the CODE II experiment was designed and conducted after the analysis of the CODE I results had been completed, it provided an opportunity to streamline the data collection and rating procedures based on the CODE I results (Deckert et al., 1992). Simplifying the materials and procedures reduced the time and effort required to produce meaningful results in CODE II. In CODE II we eliminated questionnaire items and ratings for which there was no variance in CODE I. We also shortened the tactical scenarios by eliminating the injection of new information at the end of the scenario because this portion of the scenario did not yield any interesting findings in CODE I. Because of the high reliability of the judges' expertise ratings in CODE I, we felt that having fewer judges perform the time-consuming ratings of the COA-development process would not jeopardize the reliability of the overall expertise ratings, and we reduced the number of judges providing process ratings for each subject from three to two. We also felt that sufficient data could be collected using fewer tactical situations for each subject, so we reduced the number of situations from three to two.

In the remainder of this section we describe the procedure and conduct of the CODE I and CODE II experiments. In our discussion we note the points at which the CODE II experiment differed from the CODE I experiment.

Overview of the Experimental Procedure

Prior to the experiment each subject was sent a letter introducing the experiment and a booklet describing the basic military scenario. Participants were asked to spend about an hour reading and studying the scenario, which, among other things, described the general disposition of enemy and friendly forces, concept of battle, and the past training and makeup of friendly units. A few participants failed to receive these materials, and time was allotted prior to the experiment for them to read through the booklet. There was no evidence that this placed these individuals at a disadvantage compared to the remainder of the participants.

When a subject entered the experiment setting we presented a briefing designed to provide an overview and timeline for the experiment, describe the procedure for each phase, describe what was expected of him, and explain some rules and artificialities. Participants were informed of the roles they were to assume: for all but one of the situations they assumed the role of division commander and for the other, brigade commander. We explained that one of the experimenters would play the role of the subject's immediate staff (e.g., G1 (S1), G2 (S2), G3 (S3), etc.) and subordinate commanders. Thus, any questions the subjects wished to ask of their staff or subordinate commanders were to be directed to the experimenter. After informing the subjects that the session would be videotaped but that their anonymity would be protected, we solicited their consent, and no one refused.

The subject was fitted with a wireless microphone, and the identifying materials on his uniform covered with tape. The subject was then given a copy of the first tactical situation description to study. At this time he was also directed to a 1:50,000 scale map with overlay that showed friendly and known enemy units for the specific situation, and to a 1:100,000 scale map with overlay depicting Corps forces and the enemy units opposing them germane to all situations.

These maps were fastened to the walls behind and to the side of the subject. Subjects took about 10 minutes to study the tactical situation materials. The presentations of the tactical situations were counterbalanced across subjects.

The CODE I and CODE II experiments were both conducted by two experimenters. Experimenter 1 (a military nonexpert) conducted the initial briefing and the debriefing at the end of the experiment session, and operated the video recording equipment. Experimenter 2 (with extensive Army operations and tactics knowledge) conducted all the activities within the tactical situations, including answering the subject's questions and eliciting the subject's rationale for his COA and (in the CODE I experiment) his response to the new information.

After the participant finished reading about the tactical situation, the experimenter asked the subject to give his initial reaction to the problem posed by the tactical situation. After providing his initial reaction, the subject then had an opportunity to think about the situation and ask his staff (role-played by the experimenter) questions. He was then asked to write out his "commander's concept," and messages (to staff members, subordinate commanders, and adjacent and higher echelons) stating how he wanted to deal with the tactical problem in the form of orders, requests, or notifications. On completing that task he was asked to sum up his COA verbally. He then responded to a set of predefined questions posed verbally by the experimenter about the tactical situation. A seven-item written questionnaire concerning his perceptions of the situation was administered next.

In the CODE II experiment, the questionnaire ended the experimental procedure associated with the tactical situation. In the CODE I experiment, after the subject completed the questionnaire the experimenter then presented the participant with new information about the tactical situation, and asked him whether, and if so how, he would modify his COA to account for the new information. This was followed by a questionnaire concerned with the subject's perceptions of the situation following the new information and his (possible) change of COA.

Table 1 shows a summary of the activities during each tactical situation.

Two (in CODE II) or three (in CODE I) tactical situations representing problems at the brigade- and division-command levels were presented to each subject in this manner. At the conclusion of the last tactical situation subjects completed a post-experiment questionnaire that included several biographical questions. Several questions comparing and contrasting the tactical situations were posed verbally by the experimenter and answered in like fashion. At the conclusion all subjects were debriefed and given an opportunity to ask questions about the experiment.

Subjects

Forty-six active-duty and retired Army officers served as experimental subjects. Three subjects held the rank of captain, 32 were majors, eight were lieutenant colonels or colonels, and three were general officers. As noted previously, one of the major goals of the CODE II experiment was to include higher-ranking officers in the experiment. Of the 11 subjects at or above the rank of lieutenant colonel, 10 participated in the CODE II experiment. On average the 26 subjects in the CODE I experiment had been in the service for 13.4 years, and in their present rank for 2.8 years. The 20 CODE II subjects had, on average, been in the service for 23.8 years and in their present rank for 4.1 years.

Table 1

Activities During Each Tactical Situation

1. Subject given description of tactical situation and directed to appropriate map for that tactical situation.
2. Subject asked for initial thought about tactical situation.
3. If a COA is not volunteered in #2, experimenter probes subject for his initial thinking about a COA.
4. Subject informed he has 20 minutes to ask questions, formulate a COA, and write out his concept and messages on forms provided.
5. Experimenter answers questions.
6. Subject informed that 6 minutes are left to complete writing.
7. Subject is asked to briefly summarize COA verbally.
8. Subject is asked questions about rationale for COA.
9. Subject completes Tactical Situation Questionnaire.
10. Subject given new information and directed to revised map overlay. [CODE I only]
11. Subject asked if new information will alter his COA in any way and, if so, how. [CODE I only]
12. Subject asked questions about rationale for response to new information. [CODE I only]
13. Subject completes New Information Questionnaire. [CODE I only]

Subjects worked independently and spent between two and three hours in the experimental setting. The active-duty officers agreed to participate in the experiment as part of their assigned duties; the retired officers were paid for their participation.

The Tactical Situations

To minimize the learning time and materials to be read, four tactical situations were derived from the same basic scenario. Thus, the locale, higher headquarters intent and concept of battle, and overall mission were basically the same for all of the tactical situations. The scenario and tactical situations were designed under the close supervision of a retired Army general officer. We decided to set the scenario in the Persian Gulf and model it loosely after certain occurrences in Operation Desert Storm. This allowed us to take advantage of a large amount of existing material, provided our military subjects with some semblance of realism, and provided a setting with which they would be relatively familiar. However, the subjects were cautioned on several occasions that this was in no way a reenactment of Desert Storm and that the enemy they faced was better led, more motivated, and more capable.

Prior to the actual CODE I experimental sessions, the scenario and tactical situations were pilot tested using three officers at an Army post not involved in the actual data collection, and subsequently reviewed by several officers at the School for Advanced Military Studies (SAMS). Feedback from the pilot test and review led to several changes to the scenario and tactical situations. During the post-pilot debriefing a majority of the participants commented that the tactical situations had been engaging, challenging, and reasonably realistic.

Situations A, B, and D require the participant to assume the role of division commander. In Situation C the subject's role is brigade commander. Accompanying each situation description is an appropriate 1:50,000 scale map and overlay showing friendly and suspected enemy locations. Although each of the situations was derived from the same basic scenario, they each offer a unique (nonoverlapping) tactical problem. Below is a brief overview of each situation.

Situation A. In this situation the three battalions of 1st Armored Division's 3rd Bde are fighting four enemy battalions. Weather has shut down the A-10s, but the Apaches have been attacking. The 3rd Bde commander reports to the subject (1st AD commander) that a sabkhat (dry lake bed) now made impassable by the rains lies between the current position and objectives in the northeast. The division commander must decide how to deal with the current fight, how to continue the division's movement north to the specified objective, and how to deal with the sabkhat. New Information: JSTARS reports eight enemy tanks and three helicopters moving west on a road toward our objectives.

Situation B. Third Army's objective is to destroy or capture five enemy divisions. 1st AD (the subject's command) is on the left of the corps front with an exposed left flank. Weather is bad and tac air cannot operate. 1st AD moved forward faster than the corp on its left and its 3rd Bde became engaged with a brigade of the enemy's T1 Division. Reports from airborne collectors showed they were picking up considerable radio traffic in the enemy's A1 Division — this division is on the other side of the corps boundary and on 1st AD's left flank. Weather forced these collectors to the ground and now JSTARS is down — essentially 1st AD is blind. It is brought to the division commander's attention that the A1 Division is formidable, having distinguished itself in a recent war with another country. Thus, 1st AD's commander must contend with keeping the division moving toward its assigned objectives and a potential threat on the left flank. New Information: JSTARS is back up and reports two columns of vehicles moving south-by-southwest from A1 Division, and a large increase in radio traffic is detected in the enemy's H1 Division.

Situation C. The subject is the commander of the 2nd Bde of the 1st AD, which is engaging the enemy's M1 Division on the left of the 1st AD's zone. The objectives are to destroy the M1 and move on to the S1 Division and destroy it. Speed is stressed, because S1 may try to slip away. Weather has improved and tac air is flying. Unfortunately, the 1st AD's left flank is exposed. JSTARS detects movement westward from the H1 Division on the left, threatening 1st AD's rear. H1 Division is across the corps boundary on 1st AD's left flank. The brigade commander must also contend with an impassable sabkhat, eight kilometers behind the defending enemy's M1 Division. New Information: lead elements of the brigade encounter an extensive minefield, which appears to extend from the ends of the sabkhat, almost reaching the brigade boundaries on either side.

Situation D. This situation is a modification of Situation A. Our desire was to create a more complex and difficult version of Situation A, which was perceived by the SAMS reviewers to be weaker than the other two. To this end the following changes and additions were made to A to transform it into Situation D.

• A battalion of enemy tanks slips through a seam between two friendly units and attacks a MLRS battery located in the lead bde's rear.

• A task force from the lead brigade making a flanking movement runs unexpectedly into an enemy task force.

• Air scouts report sighting a large number of enemy multiple rocket launchers between the division's current position and defined objectives to the north.

• A large number of enemy armored personnel carriers and other armored vehicles are reported across the corps boundary near a fairly good desert road leading to the area occupied by enemy being engaged by the 1 AD.

New Information: JSTARS reports eight enemy tanks accompanied by at least three helicopters, and perhaps a squad of helicopters, moving west on a road leading to the division's objective.

The first 13 subjects in the CODE I experiment responded to Situations A, B, and C, while the final 13 subjects responded to Situations B, C, and D. The 20 subjects in the CODE II experiment responded to situations B and D.

Measures

Three different kinds of measures were obtained: 1) super-expert military judges scored the presence of MCD expertise from the written products and observable (video recording) behavior of the subjects; 2) military nonexperts on the experiment team evaluated the presence of process behaviors hypothesized to be related to MCD expertise; and 3) subjects completed questionnaires providing self-report measures of certain beliefs and attitudes held by the subjects about certain aspects of the tactical situations, along with biographical information.

The judges and the judges' measures. Three retired (three- and four-star) general officers were employed as the super-expert judges. All have extensive command experience and experience training and evaluating the performance of Army commanders. Each has a reputation in the Army for being an expert tactician and leader, and has served an average of 30 years in the Army.

Prior to receiving any material to score, each judge was sent:

1. background and scenario materials,

2. the tactical situations,

3. a summary of the rating procedure, and

4. a subject question evaluation.[3]

They were specifically requested to review items 1, 2, and 3, and to complete and return item 4. We further requested that they formulate an expert's response to each of the situations, including their COA, the questions they would ask, and how they would deal with the new information. We

[3] The judges were asked to rate the criticality of a number of potential questions that subjects might ask. These questions were drawn from two sources: a list of possible questions drawn up by the military consultant who helped us develop the tactical situations, and the actual questions posed by the pilot-test subjects.

explained that their responses to the situations would provide an absolute standard against which they could evaluate the subjects' responses.

Copies of the subjects' written materials were sent to each judge. Working independently the judges rated the level of MCD expertise exhibited for each situation by the concept statement and message(s). The judges' instructions included the following:

> The rating scale you will use is a 7-point scale, with one end of the scale representing a novice rating and the other end an expert rating. Please make your ratings on an absolute, rather than a relative, standard. That is, compare each subject's response against an absolute standard of excellence, rather than against the other subjects. Establish your standard before beginning the ratings. One way to create an absolute standard is to carefully read over the description of the tactical situation and formulate what <u>you</u> think is the best COA. (It is a good idea to write it out.) Your response establishes one benchmark against which you can judge the subjects' responses. There may be other responses, different from yours, but of equal expertise, that you would also give a rating of expert, but your responses will at least provide an a priori standard. Next try to formulate some less-expert responses, down to the level that might be observed in an inexperienced, naive commander. This will help you set the other points on the rating scale.

We stipulated the order in which each judge would rate the subjects' written materials (situations were counter-balanced across judges and the order of subjects was randomized within judges). A copy of the COA rating form is given in Appendix A.

After a judge completed and returned his ratings of the written COAs, he was sent a second package containing information and materials for doing the Process Ratings, which were based on videotapes (also included) of the subjects' verbal responses to the tactical situations. The judges' instructions included the following:

> We are asking you to make three process ratings of the subject's degree of military command decisionmaking expertise.[4]
>
> First we ask you to rate the degree of expertise shown by his **Initial Reaction** to the situation, after he has read the description of the tactical situation, but before he begins the questioning/planning process. To elicit this reaction, the experimenter says, "What are your thoughts?" Depending upon the nature of the subject's response, the experimenter may prod with the question, "What are you thinking about doing?"
>
> Second, we ask you to rate the degree of expertise shown by the subject's **Decision Process**, including the questions he asks, his verbal summary of his COA, and his responses to the experimenter's questions about the tactical situation. In addition to your rating, we ask you to identify two or more positive and negative factors that influenced your rating.
>
> Third, we ask you to rate the degree of expertise shown by the subject's **Response to New Information**. This includes his response to the experimenter's inquiry about whether the New Information would cause him to modify his COA, and his response to the experimenter's questions about the New Information.
>
> The scale you will use for your ratings is a 7-point scale, with one end of the scale representing a novice rating and the other end an expert rating. As you did with the COA

---

[4] In the CODE II experiment, in which the subjects were not asked to respond to New Information about the situation, the judges only made two process ratings, and the instructions were modified accordingly.

ratings, please make your ratings on an absolute, rather than a relative, standard. That is, compare each subject's response against an absolute standard of excellence, rather than against the other subjects. Establish your standard before beginning the ratings. One way to create an absolute standard is to think about how you would respond to the situation — what your initial reaction would be, the important questions you would ask, how you would describe and explain the rationale for your COA, and how you would react to the new information. Your responses establish one benchmark against which you can judge the subjects' responses. There may be other responses, different from yours, but of equal expertise, that you would also give a rating of expert, but your responses will at least provide an a priori standard.

As with the written materials, we stipulated the order in which each judge would rate the subjects' videotapes (situations were counter-balanced across judges and the order of subjects was randomized within judges). Because of this random order and the lack of uniformity of the times taken by the subjects for each activity, we provided a tape-counter reference sheet that helped the judges locate the relevant behavior on each tape.

Prior to videotaping, the name, rank, and insignias on the uniform of each subject were covered with tape to conceal the subject's rank and identity.[5] A copy of the Process Rating Form is given in Appendix B.

After a judge completed the videotape process ratings he performed one last rating — an overall <u>expertise assessment</u> of each subject (the form is in Appendix B). The judges' instructions included the following:

> The information for rating each subject's overall level of MCD expertise is contained in five sets of materials: 1) the concept and message(s) that the subject wrote for each of the tactical situations; 2) your two ratings of these written materials; 3) your process ratings for the subject's verbal responses in the tactical situations; 4) any notes you took while viewing the tape, and 5) the videotape itself.
>
> Incorporating all this material into your judgment in whatever way you think is appropriate, please evaluate the subject's overall level of MCD expertise in the space provided on the Overall Rating Form. Then give the two factors that most strongly influenced your rating. Use the comments section to provide additional rationale for your rating.
>
> The rating scale you will be using is a 7-point scale, with one end of the scale representing a novice rating and the other end an expert rating. As you did with the individual situation ratings, please do the ratings on an absolute standard of excellence, rather than against the other subjects. Try to establish your standard before beginning the ratings.
>
> We recognize that we have given you very little guidance for doing the overall ratings. This is because we do not want to impose any a priori standard on you. When you have completed all the individual ratings, we would appreciate any general comments you can make about the factors you considered and the way you went about doing the overall ratings.

To summarize, each judge provided four (in CODE II) or five (in CODE I) ratings (on a seven-point scale) for each subject for each scenario (concept, messages, initial reaction, decision

[5] In the CODE I experiment one judge indicated he had recognized one subject. We believe that this was the only such case. In the CODE II experiment, in which there were a larger number of higher-ranking officers, the judges recognized several faces on the video tapes, but did not feel this hampered their ability to do the evaluations in an unbiased manner.

process, and, in CODE I only, response to new information), and one overall expertise assessment (also on a seven-point scale). The judges provided comments or a rationale for each of the ratings as well.

The behavioral measures. A premise of the experiment was that observers who were not military experts would be able to detect and rate behaviors exhibited by the subjects that are related to the subject's MCD expertise. To obtain these ratings experienced raters who were not military experts independently viewed the videotape of each subject and rated 16 (in CODE I) or 11 (in CODE II) specific behaviors.[6] Each rater holds a PhD in psychology, is knowledgeable about experimental methodology, and has performed observational ratings in the past. The two raters independently viewed the tactical situations and subjects in different random orders. Five items required a binary decision (e.g., yes/no, linear/contin-gent), one item (used in CODE I only) was a choice among six categories, and 10 (in CODE I) or six (in CODE II) items requested a rating on a five-point scale. A description of each of the nonexpert measures is given in Table 2, and note is made of the measures used only in the CODE I experiment. The rating forms are given in Appendix B.

[6] Data from two of the 16 measures used in the CODE I experiment (extent to which map used when explaining COA, and when in the planning period questions were asked) had no variance in the CODE I experiment, and therefore these measures were not used in the CODE II experiment. Three of the 16 CODE I measures pertain to the New Information and were not relevant in the CODE II experiment.

Table 2

Measures Based on Ratings of Behavior by Military Nonexperts*

Measure	Minimum Possible Value	Maximum Possible Value
Provide initial COA (yes = +)	0	1
Volunteer initial COA (yes = +)	0	1
Detail of initial COA (more detail = +)	1	5
Initial COA linear or contingent (contingent = +)	0	1
Flag an aspect of situation as critical (yes = +)	0	1
Extent to which used map when studying situation (more use = +)	1	5
Extent to which used map when explaining COA (more use = +) [CODE I only]	1	5
Describe when in planning period questions were asked (categorical) [CODE I only]	1	6
Degree of match between initial and final COA (higher match = +)	1	5
Extent to which final COA incorporates responses to questions (greater extent = +)	1	5
Final COA linear or contingent(contingent = +)	0	1
Extent to which COA takes account of time and event sequencing (to a greater extent = +)	1	5
Extent to which new situation was anticipated (to greater extent = +) [CODE I only]	1	5
Extent to which new situation was planned for (to greater extent = +) [CODE I only]	1	5
Extent to which COA was revised (to greater extent = +) [CODE I only]	1	5
Extent to which importance of not compromising the mission was voiced (to greater extent = +)	1	5

*In computing the average value across the two or three situations, for those measures that coded yes/no or presence/absence (e.g., did the subject flag something as critical) the score is the proportion of tactical situations in which a yes/presence response was given. For those measures based on rating scales (e.g., the extent to which the subject used the map), the score is the average for the two or three situations.

Questionnaires. Three different self-report questionnaires were employed — a copy of each is given in Appendix B. The Tactical Situation Questionnaire was completed by the subjects after development of the COA in each tactical situation. Subjects responded to seven questions about the tactical situation concerning, for example, complexity, initial and current uncertainty, how confident they were that their COA dealt with the problem, and how difficult it was to reach a COA. The subjects responded to each question on a seven-point Likert-like scale.

The New Information Questionnaire, as the name implies, focused on the subjects' opinions concerning the new information about the situation. Subjects (in the CODE I experiment only) responded to five questions about the complexity of the situation created by the new information, difficulty in formulating a response to the situation created by the new information, and the extent to which they felt their COA needed to be modified to accommodate the new information.

Finally, The Post-Experiment Questionnaire was administered at the end of the experimental session. This questionnaire solicited biographical information such as years in service, rank, years in rank, military schools attended, exercises attended, and combat experience. A total of six biographical variables were derived for analysis. In addition, the experimenter asked 10 questions orally and the subjects responded orally.

Analysis of Judges' Data

Each judge's ratings of the written materials yielded two measures of expertise: 1) concept statement and 2) messages. Similarly, ratings of the video tapes produced two or three

[7] process measures of expertise present in the subject's: 1) initial response to the tactical situation, 2) decisionmaking process, and 3) response to the new information [CODE I only]. These assessments produced a total of four or five component measures of expertise for each situation for each judge.

In the CODE I experiment, each of the three judges rated all the subjects on both the written and verbal (videotape) materials for three tactical situations, producing a total of 45 measures per subject. In the CODE II experiment, each of the three judges rated all the subjects on the written measures, but only two of the three judges rated each subject on the process measures. Furthermore, we used only two situations. Thus, for each CODE II subject we had a total of 20 measures (12 written and eight process).

To derive the measures used in our analyses, we performed several aggregations of these individual measures. We computed a mean for each of the four or five component scores across the two or three tactical situations, producing 10 or 15 (mean) variables for each subject. In a similar manner we computed a mean component score for each subject for each of the four or five component measures by averaging over the tactical situations and over the judges. We derived the average expertise rating for each judge for each subject by averaging over the four or five component scores and over the two or three tactical situations. Finally, we computed a mean over the 20 or 45 measures to yield the expertise score for each subject.

Each judge also rendered an overall expertise assessment for each subject for whom he did both written and process ratings. The expertise level, the primary measure of expertise for each subject, was derived by averaging the overall expertise assessments from the two or three judges. The expertise level and the expertise score correlate very highly ($r=.94$; $p<.01$) with each other.

Analysis of Secondary Measures

Military-nonexpert raters' data. The 16 (in CODE I) or 11 (in CODE II) ratings of subjects' behavior provided by the two military-nonexpert raters were compared for each subject for each tactical situation independently. The raters discussed any differences in the ratings and agreed on a single value for that rating item. This produced a single set of 11 or 16 rating scores for each subject, for each of the tactical situations. To produce the final set of 11 or 16 rating scores for each subject used in the analyses, the rating scores for each behavior were averaged over the tactical situations. For those ratings that entailed binary decisions, the average was based on the proportion of "yes" answers.

Questionnaire data. The seven measures derived from the Tactical Situation Questionnaire were averaged across the tactical situations, yielding seven scores per subject. Similarly, in the CODE I experiment the five measures produced by the Response to the New Information Questionnaire were also averaged across the three tactical situations, yielding five scores per subject.

Additional measures. After the subject finished studying the description of the tactical situation, the experimenter asked him for his initial reaction. We recorded the amount of time that

[7] Here and the next several paragraphs, when two alternative numbers are given, the smaller number refers to the CODE II experiment and the larger number to the CODE I experiment.

elapsed between the time the subject started responding to the experimenter and when he began describing his initial COA. We refer to that variable as "time to generate initial COA." This measure is recorded only in those cases where the subject provided an initial COA.

Following the subject's initial reaction to the situation, he was permitted to ask questions. We derived two measures from the questions asked: a count of the number of questions asked and the percentage of critical areas covered by the questions. For the latter measure, the set of critical questions provided by the expert judges in their Question Evaluation forms was employed. We grouped all of the questions deemed critical by any of the judges for each situation into categories, and gave a subject credit for a category if he asked <u>any</u> question pertaining to it.

The experimenter asked a number of open-ended questions concerning the tactical situation after the subject finished describing his COA. (The list of experimenter's question is included in Appendix B.) Two additional measures were based on the subject's responses to these questions. One was the mean number of "show stoppers" (i.e., future events that would cause him to rethink his COA) enumerated by the subject for each tactical situation. A second was whether the situation reminded the subject of an historical situation he had read about or a situation he had experienced in the past. For this measure we counted the proportion of situations that reminded the subject of a previous experience.

Results and Discussion

Because of methodological differences between the CODE I and CODE II experiments (three tactical situations and three judges in CODE I versus two tactical situations and two judges in CODE II), the reliability of the expertise ratings is discussed separately for each experiment. We view the CODE II analyses as a replication and verification of our methodology for eliciting and measuring MCD expertise. In the subsequent discussion of the distribution and the nature of expertise, our analyses are based on averages across scenarios and across judges, so we are able to merge the CODE I and CODE II data. Except for measures taken only in the CODE I experiment, we discuss the resultant analyses as one study.

Measuring Military Command Decisionmaking Expertise

Reliability Analyses

Judges' scores: CODE I. A primary premise of this investigation is that MCD expertise can be identified and reliably assessed. In the CODE I experiment, three super-expert judges were asked to independently render an overall expertise assessment for each subject based on the subject's performance on three tactical problems. The judges were not supplied with a definition of expertise; instead, we asked them to define it for themselves and then assess it for each of the subjects. If our methodological hypotheses are correct, then the three judges' overall expertise assessments should be highly correlated.

Table 3 shows the intercorrelation matrix of the three judges' expertise assessments, and the coefficient alpha (Cronbach, 1970; Nunnally, 1967) derived from the intercorrelation matrix. All of the correlations are significantly different from zero ($p<.05$) and are in the moderate to moderate-high range. The coefficient alpha, a measure of internal consistency and an important form of reliability, is .76. This is high, implying high reliability or consistency among the judges. Moreover, it has been shown (Nunnally, 1967) that the square root of coefficient alpha, which in this case is .87, represents the correlation with the true scores (expertise) could they be known. Table 3 also shows the intercorrelation among the three judges' average expertise ratings (computed from each judge's five component expertise scores) and the coefficient alpha derived from these intercorrelations. Compared to the expertise assessment correlations, the average expertise ratings for the three judges in the CODE I Experiment are correlated even higher, as is the coefficient alpha of .81.

Table 3

Intercorrelation of Expertise Assessment and Average Expertise Ratings for the Three Judges in the CODE I Experiment

EXPERTISE ASSESSMENT

	JUDGE 1	JUDGE 2
JUDGE 1	1.00	
JUDGE 2	.42	1.00
JUDGE 3	.48	.71
	COEFFICIENT ALPHA = .76	

AVERAGE EXPERTISE RATING

	JUDGE 1	JUDGE 2
JUDGE 1	1.00	
JUDGE 2	.58	1.00
JUDGE 3	.63	.73
	COEFFICIENT ALPHA = .81	

From all this we can conclude that there is an underlying quality called MCD expertise associated with individuals, that this expertise can be elicited using written materials and maps to pose a tactical problem in a laboratory setting, and that this expertise can be reliably measured by well-qualified judges. If any of these factors were not true, we would be hard pressed to explain the high agreement among the judges.

Table 4 shows the coefficient alphas derived from the intercorrelation of the judges for each of the five component measures of expertise. Again we see fairly high consistency among the judges. Each of the five component measures of expertise appears to be reliably assessed. Not only do the judges agree on the overall amount of expertise that is present for a subject, but they also agree on the amount of expertise present for the several component behaviors of MCD. These results further support the hypothesis that a concept of MCD expertise underlies Army officers' behavior, it can be observed, and expert judges can reliably agree on its relevant behaviors.

Table 4

Coefficient Alpha for Each of the Five Component Expertise Scores in the CODE I Experiment

COMPONENT	COEFFICIENT ALPHA
Rating of Concept	.77
Rating of Messages	.73
Rating of Initial Reaction	.72
Rating of Decision Process	.73
Rating of Reaction to New Information	.68

Earlier we mentioned that the correlation between expertise level and expertise score is .94, indicating that nearly 90 percent of the variability in one measure can be accounted for by the other measure. This result further attests to the reliability of the judges' expertise evaluations.

Judges score: CODE II. On the basis of the reliability analyses of the CODE I data, we concluded that the judges were highly consistent in their overall evaluations of subjects. Given this high level of consistency, we felt that it would not compromise reliability if each subject were rated by two of the three judges. Whereas the ratings of the written materials can be done reasonably rapidly, the process and overall ratings, which are based on evaluation of videotapes, are much more time-consuming. In order to streamline the expertise-evaluation procedure and test the effect of that streamlining on reliability, we asked all three judges to evaluate the written materials for all the subjects, but to evaluate only a subset of the subjects on the process and overall measures. This procedure allowed us to compare the ratings of the subjects based on three judges' assessments to those based on two judges' assessments.

In order to make optimal use of the judges' time and the resources available to us, we also limited the number of tapes evaluated by any of the judges. Focusing our efforts on the higher-ranking subjects, we evaluated the videotapes for 15 of the 20 CODE II subjects.

Using this evaluation procedure, we obtained COA and message evaluations from each of the three judges for each of the two situations for the 20 subjects in the CODE II sample. We obtained process ratings from two of the three judges on each of the two situations for 15 of the 20 subjects. Each judge did process ratings for 10 of these 15 subjects, and each pair of judges rated five subjects in common. There are no process ratings for five of the 20 subjects.

Because the overall ratings are based on a combination of the written and the process ratings, a judge could only do an overall rating for those subjects for whom he did both written and process ratings. Thus, we have overall expertise ratings for the 15 subjects for whom we have process ratings. As with the process ratings, each judge rated 10 subjects, and each pair of judges did overall ratings for five subjects in common.

Table 5 shows the intercorrelation matrix of the three judges' expertise assessment and the average coefficient alpha.[8] Because the correlations are based on a very small sample size (N=5),

[8] Because none of the subjects was rated by all three judges, it is not possible to obtain a coefficient alpha based on the intercorrelations among the three judges, as was done in the CODE I experiment. Rather, we computed the coefficient alpha for each pair of judges, and present the average of the three values here.

only correlations above .73 are significant, even at the p=.10 level. Thus, of the correlations reported in Table 5, only the correlations between judges 1 and 2 are statistically significant (although all but one are at least moderate in magnitude).

Table 5

Intercorrelation of Expertise Assessment and Average Expertise Rating for the Three Judges in the CODE II Experiment

EXPERTISE ASSESSMENT

	JUDGE 1	JUDGE 2
JUDGE 1	1.00	
JUDGE 2	.73	1.00
JUDGE 3	.25	.48
AVERAGE COEFFICIENT ALPHA = .59		

AVERAGE EXPERTISE RATING

	JUDGE 1	JUDGE 2
JUDGE 1	1.00	
JUDGE 2	.95	1.00
JUDGE 3	.56	.02
AVERAGE COEFFICIENT ALPHA = .58		

The one potentially disconcerting value in Table 5 is the complete lack of agreement for the average expertise ratings between judges 2 and 3 ($r = .02$). Because the sample size is small, the correlation coefficient can be strongly influenced by one case of polarized ratings. In this case, the two judges disagreed on one component, the initial process rating, for one subject, with one judge ranking him highest and the other judge ranking him lowest of the five subjects the two judges rated in common. This disagreement on the initial process rating is reflected in a disagreement on the average expertise ratings. If this subject is eliminated from the sample, the correlation coefficient rises to .55, which is similar to the degree of correspondence between the two judges on the overall expertise assessment, and the average coefficient alpha rises to .73, which is very similar to the alpha value reported for the CODE I study in Table 3.

Table 6 shows the coefficient alphas derived from the intercorrelation of the judges' ratings for each of the four component measures of expertise. The coefficient alphas for the ratings of concept and messages are based on the intercorrelations among the three judges, while the process ratings (initial reaction and decision process) are based on the three sets of intercorrelations among the three pairs of judges. The magnitude of the coefficient alpha values for the concept and message ratings are very similar to those shown in Table 4 for the CODE I experiment. The coefficient-alpha values for the process ratings, which are based on three groups of five subjects rather than one group of 20 subjects, are somewhat lower. Furthermore, as noted above, two judges disagreed strongly on the rating of the initial reaction for one subject. If that subject is removed from the sample, the average coefficient alpha rises from the .45 shown in Table 6 to .70, which is almost identical to the coefficient alpha reported for the CODE I study in Table 4 (alpha = .72).

Table 6

Coefficient Alpha for Each of the Four Component Expertise Scores in the CODE II Experiment

COMPONENT	COEFFICIENT ALPHA
Rating of Concept	.70
Rating of Messages	.77
Rating of Initial Reaction	.45
Rating of Decision Process	.59

The correlation between the expertise levels and expertise ratings for the 15 subjects for whom these measures are available is .94, which is the same as it is in the CODE I sample. The high correlation between the two measures of expertise and the correspondence of this value with the correlation between the two measures found in the CODE I study attest to the reliability of the judges' expertise evaluations.

The similarity of the reliability measures for the written (concept and message) components of the two studies indicates that the reduction in the number of tactical situations from three in CODE I to two in CODE II did not in any way compromise the reliability of the judges' evaluations of the written materials. It is more difficult to compare the reliability measures for the process components because the number of subjects rated in common by each pair of judges was sharply reduced. However, the magnitude of the correlations and coefficient alphas, especially given the small sample sizes, in conjunction with the consistent reliability for the written materials, supports our conclusion that MCD expertise ratings can be assessed reliably.

Military-nonexpert ratings: CODE I. After an initial calibration period but prior to the discussion and arbitration of the rating measures, the rater assessments for each of the 16 ratings were compared using coefficient alpha. The average coefficient alpha for all rating scales that had variance was .70. This average coefficient is based on the ratings from a subsample of 13 subjects over the first three tactical situations. We can see that prior to the discussion and arbitration period there was already a moderately high agreement between the two military-nonexpert raters. We therefore maintain that after arbitrating the differences existing between the two ratings, a highly reliable set of 16 ratings covering 16 aspects of a subject's command decisionmaking behavior was achieved.

Military-nonexpert ratings: CODE II. We calculated the interrater reliability using the same procedure as we did for the CODE I experiment. In the CODE II experiment the average coefficient alpha for all rating scales that had variance was .73. This value is based on the ratings for the 15 subjects for whom process and overall expertise evaluations were made, and is very similar to the coefficient alpha for inter-rater reliability in the CODE I analysis, supporting the conclusion we reached based on the CODE I data that military-nonexpert raters can reliably assess observable components of MCD expertise.

Distribution and Description of Subjects' Expertise

The analyses presented in the remainder of this section are based on a combination of the CODE I and CODE II data. These analyses are based on expertise measures and rating scores averaged across tactical situations and across judges. Because comparable average measures can be obtained in both studies, we are able to combine the data from the two studies. The only

exception is that in CODE II we did not ask for the subjects' reactions to new information, and so analyses of measures related to the new information are based on the CODE I data only.

With 26 subjects in the CODE I experiment and 20 subjects in the CODE II experiment, we have a total of 46 subjects. Because we did not obtain process and overall expertise measures for five of the 20 CODE II subjects, most of the analyses reported below are based on a sample of 41 subjects. For measures of reactions to the new information, the sample size is 26 (i.e., CODE I subjects only).

Expertise level. Figure 6 shows the frequency distribution of expertise level (the average overall expertise assessment across judges) for the 41 subjects for whom an overall expertise assessment was given. Expertise level is symmetrically distributed with a mean of 3.24 and a standard deviation of 1.20. The scores ranged from a low score of 1.33 to a high score of 5.67 (on a 7-point scale). It was our desire to obtain a sample of officers who differed in their MCD expertise and to employ a methodology that was capable of differentiating among these levels. The range of scores and the shape of the distribution suggest that our methodology was apparently able to discriminate among individuals possessing varying levels of expertise.

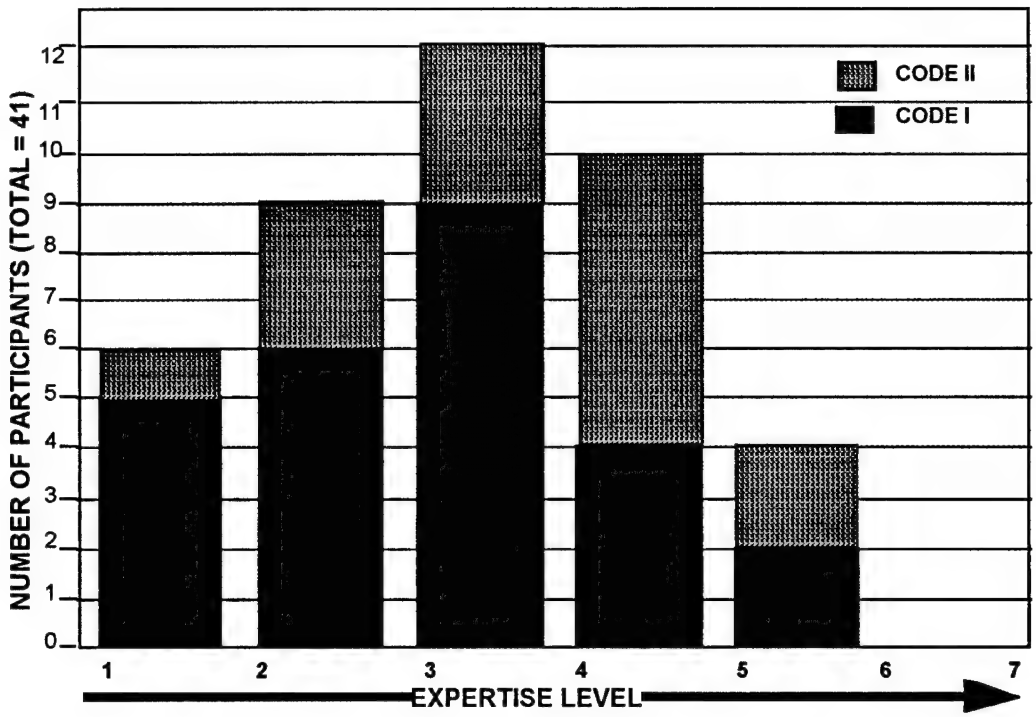


Figure 6. Distribution of expertise levels plotted as a histogram

Identifying the high- and low-expertise subject groups. To facilitate subsequent testing of our hypotheses, we used the subjects' expertise levels to identify two extreme subsets of subjects, the high- and low-expertise subject groups. Those subjects who were more than one standard deviation below the mean on expertise level were placed in the low-expertise group, and those subjects who were more than one standard deviation above the mean on expertise level were placed in the high-expertise group. Using this procedure we identified 11 low-expertise subjects and nine high-expertise subjects.

In order to insure that we had identified the extremes of our sample, we wanted to limit the number of subjects in each of the extreme groups to a maximum of 20 percent of the total sample. To achieve this goal, we also took into account the subjects' average rating score (based on the mean of the written and process ratings). In the low-expertise group, for those subjects whose expertise level was very close to one standard deviation below the mean (i.e., expertise assessment = 2.0), we eliminated from the group those subjects whose average rating score was not more than one standard deviation below the mean. In the high-expertise group, for those subjects whose expertise level was very close to one standard deviation above the mean (i.e., expertise assessment = 4.5), we eliminated from the group the subject whose average rating score was not more than one standard deviation above the mean.

This resulted in two equal-sized groups of eight subjects, representing the extremes in our sample. Table 7 compares the mean component scores and expertise levels for the two extremes.

Table 7

Comparison of High- and Low-Expertise Groups on Five Mean Component Scores and Expertise Level

	low-expertise		high-expertise			
Variable	**mean**	**st dev**	**mean**	**st dev**	**t-value**	**p**
Concept	1.8	0.3	4.0	0.5	10.4*	.000
Messages	1.5	0.2	4.0	0.5	13.7*	.000
Initial Response	2.0	0.3	4.8	0.6	7.6*	.000
Decision Process	1.9	0.7	5.0	0.4	8.1*	.000
New Information	2.0	0.2	4.6	0.4	13.4**	.000
Expertise Level	1.6	0.3	4.9	0.5	17.4*	.000

*degrees of freedom=14.
**CODE I subjects only: degrees of freedom = 6.

Background and experience of subjects. The subjects ranged in rank from captain to general. The modal ranking for the subjects was major, with 72 percent of the sample falling into this category. Time served in the Army ranged from six to 35 years, with an average of 18 years. By far the most-recent service schooling received by the subjects was the CGSC, with 83 percent of the subjects at or below the rank of lieutenant colonel reporting this as their most-recent service school attended. All the subjects at the rank of colonel reported the Army War College as their most-recent service school attended.

According to the ratings of our judges, the subjects also represent a range of expertise levels. We found no significant relationship between expertise level and the rank, time in service, or time in grade of the subjects. The correlation between rank and expertise level was .20, which is not significant (p=.21). The correlation between expertise level and time in service is .25, which is also nonsignificant (p=.14). The weak, but positive, correlation between rank and expertise level was reflected in the fact that no captains fell into our high-expertise group and no subject with a rank of colonel or higher fell into our low-expertise group. As noted above, the majority of subjects in the sample were majors, and the expertise ratings for the majors ranged from the lowest value of 1.33 to the highest value of 5.67.

Subjects did show differences in their assignment histories, however, and we noted a strong relationship between the subjects' experience and their expertise level. Specifically, officers who had served as commanders at or above the brigade level, S-3 operations or plans officers at the brigade level, or G-3 operations or plans officers at the division level were much more heavily

represented among the more-expert subjects. We rank-ordered the subjects by their expertise level, and divided them in half. Table 8 shows the relationship between expertise level and command or S-3/G-3 experience.

Table 8

Relationship Between Experience and Expertise Level

Expertise Level	**No Experience as Cmdr or**	**Experience as Cmdr or S-3/G-3**	**Total**
Bottom half of subjects	15	5	20
Top half of subjects	4	15	19
Total	20	19	39*

*Of the 46 subjects in the sample, biographical data are missing for two subjects and expertise was not rated for five subjects.

Table 8 shows that in the top half of the subjects ranked by expertise rating, 15 out of 19 had more than six months of S-3 brigade or G-3 division experience as an operations or plans officer or commander. Only five of the lower-ranking 20 subjects had such experience.[9] Subjects in the experiment may have developed greater command expertise because of their experience in S-3 or G-3 plans or ops, or it is possible that more-capable individuals have a greater tendency to be assigned to such duties. The subjects perceived their experience as helpful, however, especially their experience as plans officers, which was mentioned by several subjects as relevant to their tasks in the experiment. One of the subjects with a high expertise level commented that his experience as a brigade plans officer had helped him in the experiment because it "dealt with large formations and division-level operations."

The Nature of Expert Performance

We have shown earlier that the judges were rating something that they believed to be "MCD expertise." What is the nature of this expertise? We can gain insight into the factors contributing to MCD expertise by correlating the judges' ratings with the secondary measures of expertise suggested by our theoretical framework, by analyzing the subcomponents of the judges' ratings, and by examining the judges' explanations of their ratings.

Relationships Between Expertise Level and Secondary Measures

In order to evaluate the theoretical framework described earlier, we analyzed the correlation between a set of secondary measures of expertise based on that framework and the subjects' expertise levels. These secondary measures are of two major types: measures based on military-nonexpert ratings of the videotapes and measures based on subjects' responses to questions about the tactical situations.

Table 9 shows the correlations between the subjects' expertise levels and the secondary measures based on the military-nonexpert ratings of the videotapes. Table 10 shows the correlations between expertise level and the secondary measures based on the subjects' responses to questions about the tactical situations. The correlations reported in these two tables were computed using the 41 subjects in the sample for whom we have expertise level ratings. When

[9] Chi Square = 11.35, df = 1, $p < .001$.

the correlation is significant at p=0.10 or less, we report the significance level of the correlation in the table. When the p-value for the correlation is greater than 0.10, we simply report the correlation as nonsignificant (ns).

Table 9

Correlations Between Subjects' Expertise Level and Secondary Measures Based on Nonexpert Ratings of the Videotapes.

Measure	**Corre-lation**	**p-value***
Volunteer initial COA (yes = +)	.20	ns
Provide initial COA (yes = +)	.12	ns
Initial COA linear or contingent (contingent = +)	.33	.039
Detail of initial COA (more detail = +)	.39	.014
Time to generate initial COA	.05	ns
Degree of match between initial and final COA (higher match = +)	.14	ns
Number of questions asked	.17	ns
Percent of critical areas covered	.40	.010
Extent to which final COA incorporates responses to questions (greater degree = +)	.37	.019
Flagged an aspect of situation as critical (yes = +)	.12	ns
Extent to which used map when studying situation (more = +)	.33	.035
Extent to which final COA takes account of time and event sequencing (to a greater extent = +)	.48	.001
Extent to which importance of not compromising the mission was voiced (to greater extent = +)	.31	.046
Final COA linear or contingent (contingent = +)	.36	.021
Extent to which new situation was anticipated (to greater extent = +) [CODE I only]	.22	ns
Extent to which new situation was planned for (to greater extent = +) [CODE I only]	.29	.074
Extent to which COA was revised (to greater extent = +) [CODE I only]	.10	ns

*two-tailed p-values: n = 41 except for CODE I items only, where n = 26.

Table 10

Correlations Between Subjects' Expertise Level and the Secondary Measures of Expertise Based on the Subjects' Responses to Questions About the Tactical Situations

Measure	Corre-lation	p-value*
Situation reminds you of previous experience or historical situation (yes = +)	-.04	ns
Perceived complexity of tactical situation (more complex = +)	.39	.011
What percentage of the information you would have liked were you able to obtain (obtained high percent of information needed = +)	-.08	ns
Perceived adequacy of time allocated for COA development (adequate time = +)	-.39	.011
Number of show stoppers mentioned	.26	.100
Perceived initial uncertainty of situation (more uncertain = +)	.07	ns
Perceived final uncertainty of situation (more uncertain = +)	-.03	ns
Confidence in COA (more confidence = +)	-.16	ns
Difficulty in reaching a COA (more difficult = +)	.12	ns
Perceived complexity of situation with new information (more complex = +) [CODE I only]	.24	ns
Perceived difficulty of responding to new information (more difficult = +) [CODE I only]	.28	ns
Perceived degree to which COA modified by New Information (greater modification = +) [CODE I only]	.10	ns
Perceived uncertainty in situation with new information (greater uncertainty = +) [CODE I only]	.15	ns
Confidence that response handles new information (greater confidence = +) [CODE I only]	.15	ns

*two-tailed p-value: n = 41 except for CODE I only items, where n = 26..

More than half (9 of 17) of the correlations between expertise level and the measures based on the nonexpert ratings proved to be significant. While there was no relationship between degree of expertise and whether the subjects provided an initial COA, when they did provide one the more-expert subjects provided a more-detailed COA and one that contained evidence of contingencies. The more-expert subjects asked more critical questions, and made greater use of the responses to the questions they asked in developing their COAs. The more-expert subjects made more use of the map while they were studying the situation and expressed more concern about not compromising the mission. The final COAs prepared by more-expert subjects took timing and sequencing factors into account, and they contained evidence of contingency planning to a greater extent. When new information was received, the COAs of the more-expert subjects had already accounted for the new situation to a greater extent than those of the less-expert subjects.

We had hypothesized that the richer mental model of more-expert subjects would make them more likely to immediately volunteer an initial plan of action or to provide one when prompted. We found no empirical support for that hypothesis in the data, however. It is possible that the amount of detail provided about the situation before the subject could ask questions was not sufficient to encourage an "off-the-top" COA, or that something else in our experiment procedure discouraged an initial COA. On the other hand, when subjects did provide a COA, those provided by the more-expert subjects contained more detail and more evidence of contingency planning than did those of the less-expert subjects.

Among the 14 secondary measures based on the subjects' responses to questions about the situations, there were three significant correlations with expertise level. More-expert subjects

perceived the tactical situation as more complex, perceived the time available to develop a COA as less adequate, and enumerated more show stoppers than did the less-expert subjects.

Judges' Component Measures: Product and Process

As discussed earlier, the judges provided four (CODE II) or five (CODE I) component ratings of the expertise shown by each subject in each scenario, as well as an overall expertise assessment. Two of these five component ratings were "product quality" ratings based solely on written materials — the judges rated each written statement of concept and also the written messages prepared to accomplish the subject's COA. The judges then provided two or three expertise ratings based on their viewing of videotapes — a rating of the expertise shown by the subject's initial response to the situation, a rating of the expertise shown during the COA-development process, and, in the CODE I experiment, a rating of the expertise shown by the subject's response to the new information introduced at the end of the scenario. These three ratings were "process quality" ratings based on the behavior that the judges observed on the videotapes, including the questions asked by the subjects and the explanations given for each COA.

We expected that some of our secondary measures of expertise based on the nonexpert ratings of the videotapes and the subjects' responses to questions might be more highly correlated with the judges' product quality ratings, while others might be more highly correlated with the judges' process quality ratings. Measures that predict observable actions, such as using the map as a visualization tool, should be correlated with the process quality ratings but not necessarily with the product ratings (unless use of the map resulted in a better written concept and message set). Other measures, such as the amount of detail in the initial COA as rated by the nonexperts, should correlate with the product quality ratings, but might also correlate with the process ratings, depending on which aspects of the process the judges found most salient in preparing their ratings.

The judges' component ratings were collapsed into two ratings: a product rating that is the average of the two ratings based on written materials, and a process rating that is the average of the two or three ratings based on the videotapes. Figure 7 shows the significant correlations (with a p value of 0.10 or less) that were found between the secondary measures and the judges' product and process ratings. Not surprisingly, the judges' product and process ratings were highly correlated with each other (.79) and with their overall expertise assessments (.81 for product and .94 for process).

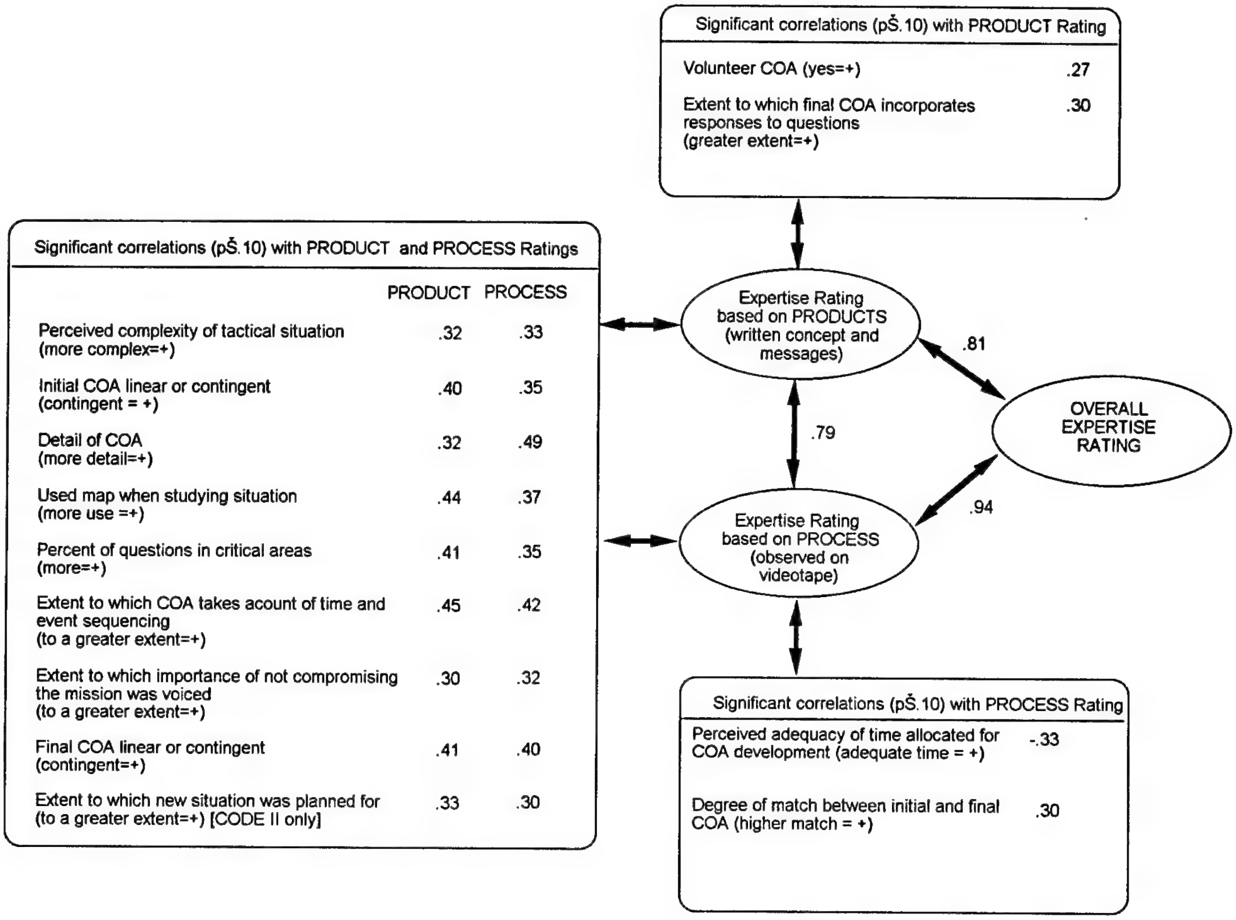


Figure 7. Significant correlations between judges' product and process ratings and secondary measures.

There were two secondary measures that were significantly correlated with the judges' product ratings but not with their process ratings. Among subjects who provided an initial COA, the more-expert subjects (as measured by the judges' rating of their written products) were more likely to volunteer an initial COA than less-expert subjects. Note that this measure is correlated with the product expertise rating, but not with the overall expertise level. The degree to which answers to questions were used to modify the COA was also significantly correlated with the product ratings.

Two secondary measures were found to be significantly correlated with the process ratings but not with the product ratings. More-expert subjects were less likely to have perceived the amount of time allocated for COA development as adequate than were the less-expert subjects. Also, there was a higher degree of match between the initial and final COAs for the more-expert subjects than there was for the less-expert subjects.

Most of the significant secondary measures were correlated with both product and process ratings. Eight nonexpert-based measures and one questionnaire-based measure were correlated with both the product and process ratings. The nonexpert-based measures included the amount of detail in the initial COA, and whether it contained evidence of contingency planning, the

percentage of critical areas covered, whether the final COA was linear or contained contingencies for possible events, the extent to which the COA took timing and event sequencing into account, and the extent to which a new situation (from the introduction of new information) was already planned for in the original COA. The extent to which the subjects expressed concern about not compromising their mission was also significantly correlated with both product and process ratings. It is interesting that the greater detail, robustness, and flexibility of the more-expert subjects' COAs were apparently visible to the judges not only from the subjects' written concept and message statements but also from observation of the COA-development process and the subjects' explanations of their COAs. The one questionnaire-based measure significantly correlated with both product and process ratings was the perceived complexity of the situation, with the more-expert subjects seeing the situation as more complex than less-expert subjects.

The finding that subjects who perceived greater complexity in the tactical situation developed better concept statements and messages is consistent with our theoretical framework. The expert builds a richer mental model of the situation than the nonexpert, and may be aware of complexities and missing information that the less-expert tactician overlooks. A situation may appear less complex to the nonexpert simply because he does not have the knowledge or experience to be aware of its true complexity. Evidence that experts build a richer mental model of the situation is also provided by the finding that more-expert subjects provide greater evidence of planning for contingent events in both their initial and their final COAs.

These findings are consistent with the idea that the expert commander's richer mental model supports more-detailed COA development and allows the expert to visualize dynamic changes in space and time, thus supporting the development of a COA that takes timing and event sequencing into account. This mental model allows the expert to develop a more flexible and robust plan that contains contingencies and can cover new situations as they occur.

Comparisons Between High- and Low-Expertise Groups

Earlier in this subsection we analyzed the correlations between secondary measures of expertise and expertise level for the entire sample of subjects in order to evaluate the nature of MCD expertise. Another way of investigating the relationship between expertise level and these secondary measures is to compare the mean ratings for the two extremes of our expertise distribution: the low- and high-expertise groups.

Table 11 shows the mean scores for the low- and high-expertise groups on the secondary measures of expertise based on the nonexpert ratings of the videotapes. Table 12 shows the mean scores for measures based on subjects' responses to questions. We used the t-test to test whether the two means are significantly different from one another. The t-value and the significance of the value are also reported in Tables 11 and 12. All the t-values were tested for significance with 14 (or 6 in the case of measures used only in the CODE I experiment) degrees of freedom. We report the exact probability for t-values significant at less than or equal to .10, and report all other t-values as ns (nonsignificant).

Table 11

Means and t-test Values for Low- and High-Expertise Groups on Secondary Measures of Expertise Based on Military-Nonexpert Ratings

Measure	Mean Scores		t-value	p-value
	Low Group (n=8)*	High Group (n=8)*		
Volunteer initial COA (yes = +)	0.4	0.6	0.86	ns
Provide initial COA (yes = +)	0.8	0.9	1.14	ns
Initial COA linear or contingent (contingent = +)	0.3	0.7	1.90	.080
Detail of initial COA (more detail = +)	1.7	2.7	3.88	.002
Time to generate initial COA (in minutes)	3.0	2.2	-0.77	ns
Degree of match between initial and final COA (higher match = +)	2.6	3.1	1.56	ns
Number of questions asked	7.4	9.3	0.88	ns
Percent of critical areas covered	43.9	67.1	2.38	.032
Extent to which final COA incorporates responses to questions (greater degree = +)	1.8	2.7	2.96	.010
Flagged an aspect of situation as critical (yes = +)	0.2	0.3	0.67	ns
Extent to which used map when studying situation (to greater extent = +)	1.8	2.7	2.13	.052
Extent to which final COA takes account of time and event sequencing (to a greater extent = +)	2.0	2.8	1.89	.054
Extent to which importance of not compromising the mission was voiced (to greater extent = +)	2.3	2.8	1.43	ns
Final COA linear or contingent (contingent = +)	0.1	0.6	3.86	.002
Extent to which new situation was anticipated (to greater extent = +)	1.9	2.1	0.37	ns
Extent to which new situation was planned for (to greater extent = +)	1.8	2.2	1.10	ns
Extent to which COA was revised (to greater extent = +)	2.1	1.9	-0.36	ns

* For the last three measures, which are based on the new information, n = 4.

Table 12

Means and t-Test Values for Low- and High-Expertise Groups on Secondary Measures of Expertise Based on Subjects' Responses

Measures	Mean Scores		t-value	p-value
	Low Group (n=8)*	High Group (n=8)*		
Situation reminds you of previous experience or historical situation (yes = +)	0.6	0.6	0.09	ns
Perceived complexity of tactical situation (more complex = +)	3.4	4.5	2.96	.010
What percentage of the information you would have liked were you able to obtain (obtained high percent of information needed = +)	61.6	56.7	-0.81	ns
Perceived adequacy of time allocated for COA development (adequate time = +)	4.3	3.5	-1.38	ns
Number of show stoppers mentioned	1.3	1.7	1.50	ns
Perceived initial uncertainty of situation (more uncertain = +)	4.3	4.8	1.37	ns
Perceived final uncertainty of situation (more uncertain = +)	3.7	4.0	0.84	ns
Confidence in COA (more confidence = +)	5.8	5.5	-0.66	ns
Difficulty in reaching a COA (more difficult = +)	2.9	3.4	0.95	ns
Perceived complexity of situation with new information (more complex = +)	2.8	3.5	0.75	ns
Perceived difficulty of responding to new information (more difficult = +)	2.4	2.8	0.47	ns
Perceived degree to which COA modified by new information (greater modification = +)	2.8	2.7	-0.09	ns
Perceived uncertainty in situation with new information (greater uncertainty = +)	4.1	4.0	-0.13	ns
Confidence that response handles new information (greater confidence = +)	5.7	5.2	0.73	ns

* For the last five measures, which are based on the new information, n = 4.

The differences between the mean military-nonexpert ratings for the low- and high-expertise groups support the correlational analysis reported earlier. Even with the relatively small sample size, almost half (seven of the 17 t-tests based on the nonexpert ratings show significant differences in the mean ratings and, in general, these differences parallel the significant correlations reported previously. In this subsection we only note cases where the two types of analyses (correlation and t-tests) produce somewhat different results.

The major difference between the results based on the correlational analysis and the results based on the t-tests is that whereas the correlational analysis showed a significant relationship between expertise level and the extent to which subjects voiced concerns about not compromising the mission, there was not a significant difference between the high- and low-expertise groups on this variable. Although the mean for the high-expertise group was higher than the mean for the low-expertise group, the difference did not reach an acceptable level of significance ($p = .17$).

Similarly, for the measures associated with the subjects' reactions to new information about the situation (presented only in the CODE I experiment), the analysis of the means did not provide support for the finding based on the correlational analysis that more-expert subjects planned for the new situation to a greater extent than did the less-expert subjects.

For the secondary measures based on the subjects' responses, the analysis of differences between means for the low- and high-expertise groups in Table 12 shows that of the three measures found to be significant in the correlational analysis, only the perceived complexity is significant in the means analysis. The number of show stoppers and perceived adequacy of time are in a direction consistent with the correlations but fall somewhat short of significance.

Neither subjects' perceived initial uncertainty nor their final uncertainty in a situation was found to be significantly different in the t-tests of the high- versus low-expertise groups. Table 12 does indicate that subjects in both groups reduced their perceived uncertainty as they asked questions and developed a COA, however. The mean uncertainty of the low-expertise subjects decreased from 4.3 to 3.7 ($t=1.89$, $df=7$, $p=.101$), while the mean uncertainty of the high-expertise subjects decreased from 4.8 to 4.0 ($t=2.40$, $df=7$, $p=.048$). The somewhat larger decrease in uncertainty in the high-expertise group suggests that while both groups acted to reduce their uncertainty, the more-expert subjects may have done so more effectively.

The overall degree of correspondence between the results obtained from the correlational analysis performed on the whole sample and the analysis of the mean differences between the low- and high-expertise groups provides support for the empirical findings relating secondary measures of expertise to overall expertise levels.

Summary of Secondary Measures

Table 13 summarizes the results of the four analyses conducted to relate the secondary measures to expertise: 1) the correlation of secondary measures to overall expertise level for the entire subject population (n=41), 2) the correlation of the secondary measures with an expertise rating based only on written products (n=46), 3) the correlation of the secondary measures with an expertise rating based only on process (n=41), and 4) t-tests comparing the means of the secondary measures for the low- and high-expertise groups (n=16). The table shows, for each secondary measure, whether it was found to be significantly related to expertise in each of the four analyses.

Table 13

Summary of Significant Relationships of Secondary Measures to Expertise Level

Measure	Expertise Level	Product Expertise	Process Expertise	t-test
Volunteer initial COA (yes = +)		X		
Provide initial COA (yes = +)				
Initial COA linear or contingent (contingent = +)	X	X	X	X
Detail of initial COA (more detail = +)	X	X	X	X
Time to generate initial COA				
Degree of match between initial and final COA (higher match = +)			X	
Number of questions asked				
Percent of questions in critical areas	X	X	X	X
Extent to which final COA incorporates responses to questions (greater degree = +)	X	X		X
Flagged an aspect of situation as critical (yes = +)				
Extent to which used map when studying situation (to greater extent = +)	X	X	X	X
Extent to which final COA takes account of time and event sequencing (to a greater extent = +)	X	X	X	X
Extent to which importance of not compromising the mission was voiced (to greater extent = +)	X	X	X	
Final COA linear or contingent (contingent = +)	X	X	X	X
Extent to which new situation was anticipated (to greater extent = +)				
Extent to which new situation was planned for (to greater extent = +)	X	X	X	
Extent to which COA was revised (to greater extent = +)				
Situation reminds you of previous experience or historical situation (yes = +)				

Table 13 (Continued)

Summary of Significant Relationships of Secondary Measures to Expertise Level

Perceived complexity of tactical situation (more complex = +)	X	X	X	X
What percentage of the information you would have liked were you able to obtain (obtained high percent of information needed = +)				
Perceived adequacy of time allocated for COA development (adequate time = +)	X		X	
Number of show stoppers mentioned	X			
Perceived initial uncertainty of situation (more uncertain = +)				
Perceived final uncertainty of situation (more uncertain = +)				
Confidence in COA (more confidence = +)				
Difficulty in reaching a COA (more difficult = +)				
Perceived complexity of situation with new information (more complex = +)				
Perceived difficulty of responding to new information (more difficult = +)				
Perceived degree to which COA modified by new information (greater modification = +)				
Perceived uncertainty in situation with new information (greater uncertainty = +)				
Confidence that response handles new information (greater confidence = +)				

Three measures were significantly related to expertise in three of the four analyses: the extent to which the COA was modified based on responses to questions, the extent to which the importance of not compromising the mission was voiced, and the extent to which the new situation was planned for in the COA.

The perceived adequacy of time allocated for COA development was significant in two of the four analyses. Three measures were found to be significant in only one analysis: the proportion of times an initial COA was volunteered, the degree of match between the initial and final COA, and the number of showstoppers mentioned. In the case of these three measures that correlated with only one measure of expertise, the correlations were weak and the significance level was only in the $.05<p<.10$ range, suggesting that these may not be stable relationships.

Overall, the secondary measures based on the military-nonexperts' ratings of the videotapes showed somewhat stronger and more consistent relationships to expertise level than the measures based on the subjects' responses to questions. Of the 17 nonexpert-based measures used, 11 were significant in at least one analysis, and six were significant in all four analyses. Just three of the 14 question-based measures were significant in at least one analysis, and only one was significant in all four.

Judge's Comments on the Low- and High-Expertise Subjects

The judges had multiple opportunities to rate each of the subjects in each of the scenarios — first based on the quality of their written concept statements and their accompanying written messages, and then based on the viewing of videotapes of the COA-development process. For each of the five component ratings given for each subject in each scenario, the judges were asked to describe the basis for their judgment.

The judges' comments can provide insight into the completeness of the set of secondary measures that were based on the theoretical framework and on the adequacy of that framework. To what extent did the judges volunteer the same factors in their comments that were included in the secondary measures? Were there other factors that were important to the judges in making their ratings that were not predicted by the framework?

Although there was considerable variation among the judges in the focus of their comments, and variation from subject to subject in the positive and negative qualities cited by the judges as the basis for the component ratings at each stage, certain consistent themes emerged from the judges' written comments. This is most clearly seen by comparing the judges' comments on the eight high-expertise and eight low-expertise subjects. The points cited by the judges to explain why they rated these subjects as high or low provide insight into the judges' underlying concepts of MCD expertise.

The Low-Expertise Subjects. The judges' most-frequent criticism of the low-expertise subjects was that their plans and their planning process lacked sufficient substance and detail. One plan was described as "flimsy," another as "tentative," and one judge commented that "everything is missing." "Too general" and "no specifics" were frequent comments. Among the critical issues and details that the judges frequently noted as missing were:

- fire support plan, including priorities for fire support
- clear designation of main attack
- sufficient maneuver guidance for brigades
- plan for use of attack helicopters and close air support
- plan for (or even consideration of) deep battle
- plan for protection of flanks
- consideration of logistics
- provision for a reserve
- use of reconnaissance

Some of the judges' comments concerned specific aspects of the subjects' plans that they considered unworkable or ill-advised (e.g., "poor way to use fires," "poor choice of fighting formation"). These specific criticisms were less frequent, however, than the comments indicating that the entire plan lacked sufficient detail.

The judges' comment that the plans of the low-expertise subjects lacked details corresponds to the secondary measure "detail of initial COA." This measure was significantly related to

expertise level in all four of the analyses conducted, as might be expected given how often the lack of details was mentioned by the judges in explaining their ratings.

The low-expertise subjects were also criticized for not understanding the mission (and their own part in it), for losing their focus on the mission (often by overreacting to some aspects of the situation), and for developing a plan with the "wrong end state." The low-expertise subjects had difficulty in understanding the boundaries for their own mission, the resources under their control, and the assistance possibly available from other units or from higher HQ. One judge pointed out that a subject was "using resources not under his control to fight an enemy outside his zone." Related comments included "didn't know who he was," and "fighting the other brigade's battle." A related criticism was that the subjects failed to ask higher HQ for the assets that they needed.

This aspect of the judges' comments is most closely related to the secondary measure "extent to which importance of not compromising mission was voiced." This measure was significantly related to expertise in three of the four analyses, consistent with the importance placed on mission focus in the judges' comments.

The judges commented that the subjects had a "poor read of the battlefield", got "down in the weeds" and could not see the big picture, or "could not read the situation" and on their failure to use METT-T in analyzing the situation. Subjects were particularly likely to ignore the time dimension in METT-T. The subjects "should have asked more questions" or, if they asked questions, "should have asked the right ones" before developing their COA, and they sometimes "wasted time on issues with no bearing on the situation." The judges also felt that the subjects did not sufficiently explore alternative COAs before settling on one.

These comments correspond to a number of secondary measures: "extent to which used map when studying situation," "percent of critical areas covered in questions asked," "extent to which final COA incorporates responses to questions," "number of show stoppers considered in planning," and "extent to which final COA takes account of time and event sequencing."

The judges also commented that subjects did not know how to properly use the weight and mass of the division, and were "fighting one brigade at a time" rather than using the whole division in an effective manner. The subjects' goals were sometimes unrealistic, leading to "not enough attention or resources to the fight at hand." "Lost focus" and "lost momentum" were other frequent comments. A common complaint was that subjects did not know how to bring together the various elements under their control in an effective way, in time as well as in space, e.g., "does not know how to put fire and maneuver together to achieve decisive combat power."

These comments are also related to the secondary measure "extent to which final COA takes account of time and event sequencing." This measure was significantly related to expertise in all four analyses, consistent with the frequency with which related concerns were mentioned by the judges.

A number of the judges' criticisms concerned the clarity of the subjects' concept statements, messages, and explanations of their COAs. Judges often found the written statements and messages unclear and hard to understand, and commented on the use of nonstandard language which they felt would be a source of confusion in a real situation. A frequent comment (based on a comparison of the written statements with the videotaped explanations) was "has trouble putting thoughts into writing."

These comments are not addressed by the set of secondary measures used in the experiment. The clarity with which thoughts are expressed is a component of expertise that falls outside the current theoretical framework, although lack of clarity may be one manifestation of lack of

detailed thinking about the plan or lack of understanding of how the plan fits together with the overall mission.

The judges occasionally gave their opinions about the source and best remedy for the subject's lack of expertise. Some of these comments suggested that the subjects required further training in a specific body of knowledge, e.g., "good ideas, but no doctrine or tactics base — needs more training," "little tactical knowledge," "does not understand mission analysis and COA-development techniques," and "does not know and use Army's battle operating systems," while others suggested that a basic talent or perspective was missing, e.g., "no feel for warfighting," "does not have good warfighter instincts," "too cautious," and "too focused on avoiding defeat rather than on defeating the enemy." Other, less-frequent comments dealt with the inability to apply abstract knowledge to specific situations, e.g., "understands AirLand Battle doctrine but cannot relate to how it unfolds on the battlefield" or "knows fundamentals of warfighting, but has great difficulty in putting that knowledge into action."

The High-Expertise Subjects. As might be expected, many of the judges' positive comments about the high-expertise subjects were mirror images of their negative comments about the low-expertise subjects. The judges frequently mentioned the presence of solid details in the plans made by the high-expertise subjects, e.g., "good priority of fires concept," "good maneuver plan," "addresses both fires and maneuvers," "concern for logistics," "provided for reserves," "contingency for flank protection," and "clarified main attack priorities." The judges pointed out that the high-expertise subjects were able to consider many factors and do many things in a short period of time. They noted that the high-expertise subjects' plans were more complete ("considered all battle operating systems") and more balanced ("balanced consideration of all battle operating systems").

The plans developed by the high-expertise subjects were far from perfect in the eyes of the judges, however, and received some of the same criticisms as those of the low-expertise subjects. Typically, the judges mentioned the strong aspects of the plans but also made criticisms concerning missing ("no concern with logistics" or "no attention to deep battle"), incomplete ("not much concern over terrain and obstacles"), unclear ("maneuvers unclear"), or incorrect ("poor use of attack helicopters") elements. The high-expertise subjects apparently received higher ratings from the judges not because their plans were complete or perfect, but because their plans included more critical details and dealt with more of the important issues than those of the low-expertise subjects.

The high-expertise subjects received positive comments for their ability to maintain focus on the mission, their reading of the situation, their ability to "see" the battlefield, their use of METT-T (especially the time element) to organize their thoughts, their asking good questions ("could get to the heart of the situation"), their examining alternative COAs, and their considering possible enemy actions. The high-expertise subjects "appreciated time/space factors, both friendly and enemy" and "understood time/distance correlation." In contrast to the low-expertise subjects, the high-expertise subjects understood "how to mass combat power," were "using mass to hasten fight outcomes," and knew how to keep focus on "momentum and mass toward objective." They were able to strike a "good balance between fighting and supporting the fight."

The judges noted that the high-expertise subjects understood their own role vis-a-vis that of corps. They focused on and understood the corps commander's intent, "understood that corps mission is primary," and "oriented on the corps mission." The high-expertise subjects also had a clearer grasp of their own mission, their own assets, the limits on their responsibilities, and the role of higher authority. They "knew what to expect from corps," made "good requests to higher HQ," and had a "good understanding of Corps CG's role in all this."

The high-expertise subjects were also clearer in their written concept statements, written messages, and verbal explanations. One judge commented that a subject's written concept was a "good visualization of how he wants the fight to go." Another said that the "messages are clear and support the intent." The judges still found room for improvement in the clarity of some of the written statements, however, even for the high-expertise subjects.

The high-expertise subjects appeared to assess risk in a way that met with the judges' approval, in contrast to the low-expertise subjects who were sometimes considered too cautious. For example, one judge commented that a subject's decision was "a risky call but the right one given his mission." Other comments included "good risk assessment" and "a good risk taker — not a gambler."

The secondary measures of uncertainty and risk assessment used in the experiment did not prove to be very meaningful. The theoretical framework suggests that expert and nonexpert tacticians differ in their responses to uncertainty and their assessment of risk, but the evidence from the experiment is indirect. To the extent that a pattern was found, it suggests that the expert may perceive more uncertainty in an initial situation than a nonexpert. The expert then either acts to reduce that uncertainty, or acts in spite of the uncertainty, or both.

Some of the judges' overall comments on the high-expertise subjects concerned their grasp of specific knowledge, e.g., "excellent knowledge of brigade and division warfighting tactics," "appreciates terrain," "knows his weapon capabilities," "understands how to fight," and "[has] outstanding knowledge of both enemy and US doctrine and tactics." Other comments dealt with talent, perspective, or attitude, e.g., "confident and assured tactically," "excellent warfighter instincts," "sound tactician with excellent instincts," and "quick and intuitive grasp of warfighting."

Many of the judges' overall comments on the high-expertise subjects indicated that while these subjects demonstrated considerable proficiency, there was room for improvement in their performance through more experience. For example, "this subject, with time and experience, will make a superb senior commander," "will be a fine analytical thinker with more experience," and "great promise here." The judges differentiated between the completeness of the subjects' plans and the speed with which they were able to develop those plans, which was expected to improve with experience. For example, "a good military thinker, but a bit too methodical for now — with experience, will be able to focus on mission and situation more quickly."

Summary of Experiment Results

The first major set of results of the CODE experiments concerns the success of the methodology. We were able to create tactical situations, using easily portable written materials and maps, that elicited MCD behavior across a range of expertise levels according to the judgment of a group of MCD super-experts. Furthermore, these judges were remarkably consistent in their ratings of the expertise of each individual subject, and in their ratings of an individual's expertise based on the evaluation of written materials and on the observation of behavior on videotape. The situations presented to the subjects and the behaviors that were elicited during the experiment (e.g., asking questions, developing a COA, explaining the COA, responding to new information) provided enough information for the judges to produce a sensitive and stable differentiation of the 46 subjects along a scale of MCD expertise. The relatively low relationship between the judges' assessment of MCD expertise and the subjects' rank and years of service indicates that the judges' ratings were not biased by the apparent age and inferred rank of the subjects (i.e., the judges did not automatically rate subjects higher on expertise just because they looked more mature and therefore more experienced). The similarity

of the CODE I and CODE II inter-judge reliability analyses also supports the soundness of our assessment methodology.

The second major set of results concerns the correlation of theory-based secondary measures of MCD expertise with the expertise ratings of super-expert judges. The correlation of measures derived from the theoretical framework with expertise level as rated by the judges increases our confidence that the framework is an accurate and useful description of MCD expertise. The comments made by the judges in association with their ratings are also useful in evaluating the relevance and completeness of the framework.

Figure 8 shows the secondary measures used in the experiment, their expected linkages to the components of expertise as described in the theoretical framework, and the direction of the actual relationships that were found between the secondary measures and MCD expertise level as rated by the judges. The secondary measures were of two major types: measures based on ratings made by military nonexperts viewing videotapes of the subjects, and measures based on the subjects' responses to a written questionnaire and to a standard set of verbal questions asked at the end of each scenario. Figure 8 shows which of the measures used in the experiment proved to be significantly related to expertise (shown in highlighted boxes with the direction of the relationship indicated) and which measures did not prove to be significantly related to expertise (shown in italics inside dashed-line boxes). A measure is shown as significantly related to expertise if it was significant in any one of the four analyses conducted (see Table 13).

The predictions of the framework for which we found the least support in the experiment involved the expert's initial assessment of the situation and retrieval of a relevant schema and associated COA from a memory store of specific experiences. This is the part of the framework that is most heavily based on other cognitive science research, specifically Klein's observation of command and control decisionmaking in naturalistic settings and Schank's theory of case-based reasoning (Riesbeck and Schank, 1989). We did not find a significant relationship between expertise and the perceived similarity of the tactical situation to previous experiences, the providing of an initial COA, or the speed with which the subjects generated an initial COA. For subjects who did provide an initial COA, there was some evidence of a relationship between expertise and the unelicited volunteering of the initial COA and the extent to which that initial COA remained in place as measured by the similarity between the initial COA and the final COA, but even this evidence was weak.

We did find consistent evidence that when subjects provided an initial COA, the high-expertise subjects provided a more-detailed COA with more contingencies than did the low-expertise subjects. The level of detail of the COAs was frequently cited by the judges in explaining their positive and negative ratings of subjects. Lack of relevant substance and detail was one of the most frequent criticisms of the low-expertise subjects, and even the high-expertise subjects were sometime criticized for failing to provide enough detail on aspects of the plan such as maneuvers or use of artillery. The finding that the high-expertise subjects' initial COAs contained more detail is consistent with the idea the expert is able to draw on previous experience to generate a more-complete schema for the tactical situation and therefore can produce a more-detailed plan for action early in the planning process.

Based on observation of the videotapes, it appears that the design and methodology of the experiment may be responsible for the failure to find more evidence of a rapidly generated initial schema and COA. We expected, based on the behavior of experts during the Phase I interviews, that the more-expert subjects might volunteer their early thoughts on a possible COA as soon as they were presented with the tactical situation. If the subject did not volunteer a COA, the experimenter was instructed to probe to see if the subject had a course of action in mind after first seeing the situation.

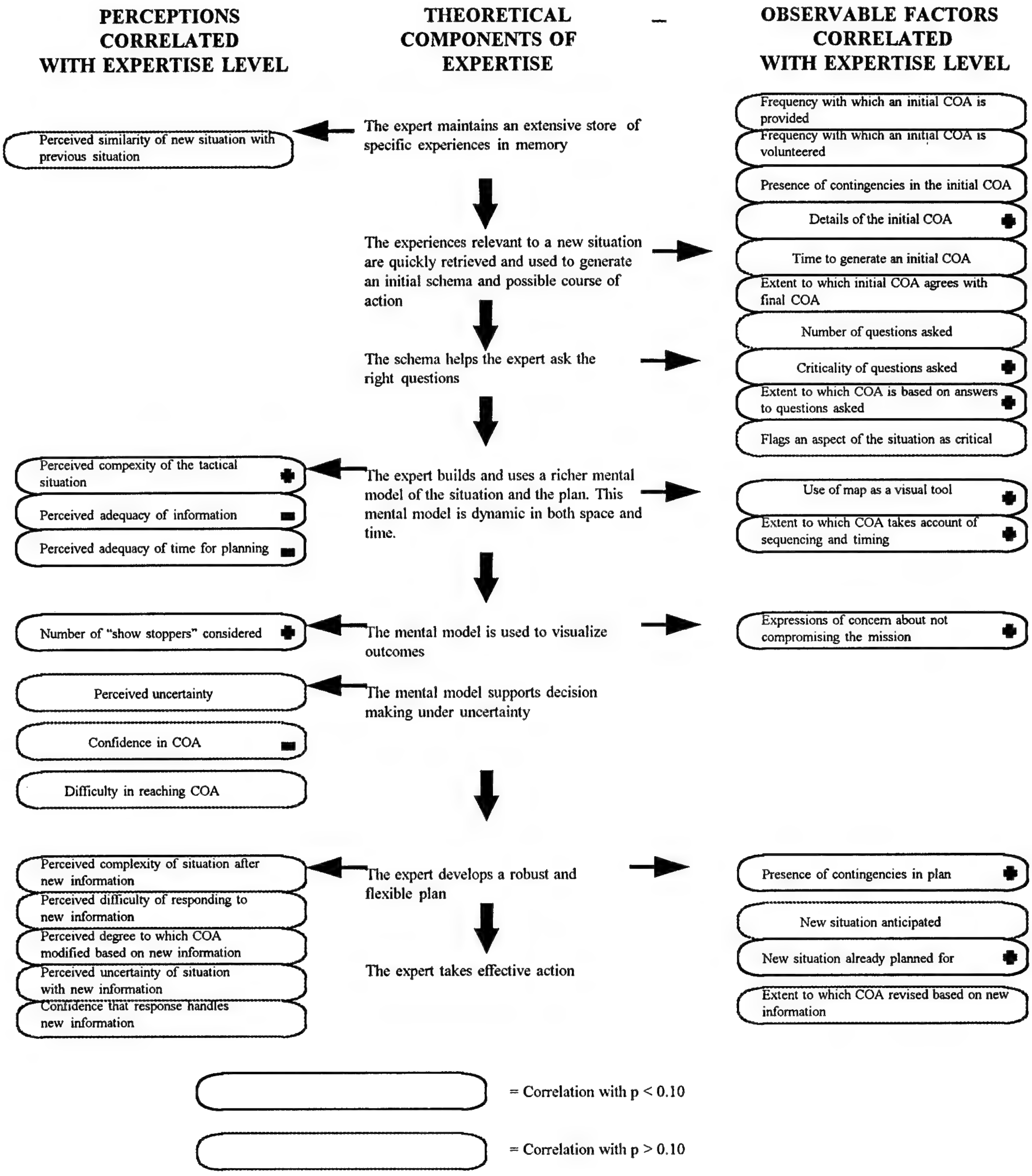


Figure 8. Secondary measures found to be correlated with expertise.

In the conduct of the experiment we found that almost all the subjects immediately wanted more information and started to ask questions rather than volunteering thoughts about a possible COA. This may be a result of the limited amount of information that was provided to the subjects at the start of the scenario. Other studies that have found evidence of rapid RPD, such as Klein's (1988) study, examined the decision process of subjects in a more information-rich real-world environment. It is also possible that the subjects' questions were driven by their early concepts of the possible COAs open to them, but that our measures were not sensitive enough to detect this link. This possibility is supported by the finding of a relationship between expertise and the asking of critical questions, but not between expertise and the volume of questions.

Another possibility is that experts do not necessarily generate a COA <u>more</u> <u>rapidly</u> than nonexperts – they <u>rapidly</u> <u>generate</u> a <u>better</u> COA. Our results indicate that both the more-expert and less-expert subjects could quickly generate a COA, but the COAs of the more-expert subjects had more detail.

Most of the remainder of the framework is well supported by the correlations found between the secondary measures and expertise levels. We found that more-expert subjects asked questions that covered a higher percentage of critical areas than less-expert subjects, and that the COAs of the more-expert subjects were based to a greater extent on the responses to their questions. This is consistent with the prediction from the framework that the expert quickly generates a schema that helps him to organize information, identify gaps in his information, and ask the right questions to fill those gaps. The judges also mentioned asking the right questions as a positive factor in their expertise ratings.

The evidence that an expert builds and use a "richer" mental model, as predicted by the framework, is diverse and indirect, but extensive. The more-expert subjects' COAs took complexities of timing and sequencing into account to a greater extent (based on nonexpert ratings) than those of the less-expert subjects. The importance of considering timing and coordination factors in the plan was often mentioned by the judges, who praised the high-expertise subjects for considering these factors. Based on this finding, and on the finding that high-expertise subjects provided more detail in both their initial and final COAs, it seems clear that the expert is able to generate and use more detail during planning than the nonexpert. The construction and use of a richer mental model is one way to explain this difference in ability.

We also found that the more-expert subjects had different perceptions about the complexity of the tactical situation, and the adequacy of the time and information provided for COA development. The more-expert subjects assessed the situation as more complex than the less-expert subjects, and rated the time and information available as less adequate. Since all subjects received identical initial situation descriptions, the experts' perceptions of greater situational complexity must have been generated from their own individual knowledge of the potentially important factors, consistent with the idea that experts draw on their experience to develop a richer mental model of a tactical situation.

Major support for the idea that more-expert subjects use their richer mental models to visualize outcomes when developing a COA comes from the frequency with which subjects verbally expressed their concern about not compromising the mission when discussing their COAs and from the evidence of contingency planning expressed in their COAs. The more-expert subjects seemed to have been fixed more firmly on the "end state" of their planned actions. They were more likely to express concern about whether any actions they might take could compromise the overall mission, indicating that they were seeing the implications of possible COAs for ultimate mission success. This is also reflected in the judges' comments. A frequent criticism of the less-expert subjects was that they lost sight of the end state or became distracted from their primary mission. More-expert subjects also mentioned alternative plans in their COAs, indicating that they had done a more thorough job than the less-expert subjects in thinking through how

their plans might play out and how they would deal with unexpected events that could jeopardize the mission.

We also observed that the more-expert subjects spent more time studying the map while they were seeking to understand the situation, again consistent with the idea that they are developing and using a mental model to visualize the situation and the consequences of possible actions. This is supported by the judges' frequent mentions of the ability to "read the battlefield" or "read the situation" in explaining their ratings.

We expected, based on the framework, that more-expert and less-expert subjects would differ in their perceptions of the uncertainty of the tactical situation, but we did not have a strong prediction about the direction of that relationship. We believed that MCD experts act more effectively under uncertainty, but we did not know whether this is because they perceive less initial uncertainty, take more actions to reduce that uncertainty, or come up with plans that are better at taking uncertainty into account.

Neither of the two questions that dealt with uncertainty was directly related to expertise. We did find a somewhat larger decrease in reported uncertainty over time for more-expert subjects, however. One possible conclusion is that experts probably do not perceive less uncertainty in tactical situations, but rather learn to live with uncertainty in their planning process and to take action even under uncertain conditions. The evidence that more-expert subjects incorporate more contingency planning into their COAs than the less-expert subjects supports this conclusion. FM-105 discusses the inevitability of "accepting risk" at some locations on the battlefield in order to achieve sufficient force and exploit success elsewhere. The judges also mentioned the ability to assess risk as a characteristic of the expert, and sometimes criticized the low-expertise subjects for being overly cautious in their risk assessment.

A second possibility is that the more-expert subjects perceived more initial uncertainty in the situation, but that by asking critical questions and dealing with the uncertainty in their plans they <u>reduced</u> their perceived level of uncertainty, and once they finished thinking about and discussing the tactical situation there was no difference between the uncertainty perceived by the more- and less-expert subjects. Because we asked the subjects about their initial uncertainty retrospectively, they may have become less attuned to their initial level of uncertainty, and responded only on the basis of their current (reduced) degree of uncertainty. A more-sensitive method of assessing their initial level of uncertainty would have been to ask the question immediately after they finished reading the tactical situation, before they asked questions or responded in any way as to how they would deal with the situation.

The framework suggests that the expert uses his mental model of the situation to visualize outcomes in order to develop a more-robust and flexible plan. We found that two measures of the robustness and flexibility of the plan were significantly correlated with expertise level: the presence of contingencies in the plan, as rated by nonexperts based on the videotapes; and the extent to which the new information introduced at the end of each scenario had already been planned for in the COA, also based on nonexpert ratings. Recall that neither of the nonexperts rating the videotapes had military training or experience, but they were able to detect the presence of contingencies in the COAs as explained by the subjects, and to assess whether the original COA covered the new situation. We did not see evidence that the more-expert subjects had anticipated the specific new situation that was introduced but, nevertheless, their plan provided for it. We also did not see evidence that the extent to which the COA was changed in response to the new information was related to expertise level, but this may be because few subjects changed their COAs in response to the new information so that there was little variation on this measure.

The finding that the more-expert subjects' plans were more likely to contain contingencies and more likely to already cover a new situation is consistent with the finding that experts

perceive more complexity in the situation and build contingencies into their planning. Because the expert sees more complexity and has a deeper understanding of how various risk factors might affect the mission, he develops a plan that is more robust to uncertainties and thus more flexible in handling new events.

None of the question-based measures that dealt with the subjects' response to the new information provided after the COA had been developed proved to be significantly related to expertise level. On the basis of the subjects' responses and the judges' comments, we believe that the new information that was introduced in the scenarios was generally not perceived as significantly changing the tactical situation. The subjects' subjective responses to this new information were therefore not a good indication of their expertise level. One reason that we eliminated this aspect of the experiment procedure from the CODE II experiment was our conclusion that the new information was not robust enough to elicit the desired responses.

Although there were a number of significant correlations between the secondary measures and expertise level, the magnitude of most of these correlations was only moderately high. The strongest measure, the extent to which the COA takes account of time and event sequencing, had a correlation of 0.48 with expertise level, and so would account for about 23 percent of the variability in expertise level. On the other hand, although the relationship between each measure taken by itself and MCD expertise was only moderately high, a combination of these measures may provide a good predictor of expertise.

Conclusions and Recommendations

Summary of Phase II Work

Toward a Theory of Military Command Decisionmaking Expertise

As indicated in Figure 5, we derived hypotheses from several aspects of our theoretical framework that we felt should be measurable in the context of the CODE experiments. The experimental results supported the following aspects of the framework:

- the expert asks the right questions and the answers to these questions influence his plan,
- the expert has a richer mental model of the situation and the plan, and this model is dynamic in both time and space,
- the expert uses his mental model to visualize outcomes in order to refine his plan, and
- the expert develops a robust and flexible plan, anticipating critical potential events and accommodating them in the form of contingencies.

The experiment, however, produced only indirect support for the following aspect of our framework:

- the expert maintains an extensive store of specific experiences in memory, retrieves experiences relevant to a new situation quickly, and uses those experiences to generate an initial schema and plan.

This aspect of the theory could be manifested in any of the following ways: 1) experts produce a better initial COA and produce it more rapidly than nonexperts, 2) experts and nonexperts produce initial COAs of equal quality, but experts produce an initial COA faster than nonexperts or 3) experts and nonexperts take the same amount of time to produce an initial COA, but the experts produce a better initial COA. The first hypothesis is the strongest: experts are both better and faster at producing an initial COA. We found support in CODE only for the third hypothesis, however. There was no evidence that experts produced an initial COA more rapidly than nonexperts, but the initial COA the more-expert subjects produced was more detailed and contained more contingencies, indicating that, based on the initial situation, experts were able to rapidly summon more relevant information than were nonexperts.

While we felt that more- versus less-expert subjects would perceive and react to uncertainty differently, the theoretical framework did not clearly indicate what direction that difference would take. In fact, we found no relationship between expertise and perceived uncertainty as reported by the subjects. We suspect that our retrospective method of evaluating uncertainty may not have been sensitive to differences between the perceived initial uncertainty of experts and nonexperts. We did find some evidence that experts act more effectively to reduce their uncertainty. Perhaps the expert is calibrated to the actual uncertainty in a tactical situation arising from all of the elements of the "fog of war." Thus, the expert may initially perceive the same situation as equally or even more uncertain than the nonexpert because of the expert's recognition of the many unknowns. The expert is accustomed, however, to planning in the presence of an acceptable level of uncertainty, and he reduces the initial uncertainty to the acceptable level by asking the right questions (as discussed above). Finally, because of his richer, more-realistic model of the situation and his plan, the expert is able to visualize outcomes and possible impediments more clearly, and therefore builds contingencies into his plan in order to deal with uncertainty. The

expert is accustomed to taking an acceptable amount of risk, however, to achieve a tactical advantage and his plan is sufficiently robust that the inherent risk does not jeopardize the mission. We suggest that the more-expert subjects immediately begin to reduce and cope with uncertainty by asking critical questions and building contingencies into their plans. Because the high-level expert deals immediately and effectively with uncertainty, we may see little difference between the uncertainty levels of experts and nonexperts.

Evaluation and Measurability of MCD Expertise

The success of the methodology clearly indicates that MCD expertise can be measured in a controlled, experimental environment. We were able to create a tactical situation, using easily portable written materials and maps, that elicited MCD behavior across a range of expertise levels according to the judgment of three MCD super experts. Furthermore, these judges were remarkably consistent in their ratings of the expertise of each individual subject, and in their ratings of an individual's expertise based on the evaluation of written materials and on the observation of behavior on videotape. The situations presented to the subjects and the behaviors that were elicited during the experiment (e.g., asking questions, developing a COA, explaining the COA, responding to new information) provided enough information for the judges to produce a sensitive and stable differentiation of the 46 subjects along a scale of MCD expertise even though the rank and years of service of the majority of the subjects did not differ greatly.

Many of the secondary measures, which could be observed by military nonexperts, were correlated with expertise level as rated by the super-experts. They did not, however, account for a sufficient portion of the variance to be, by themselves, reliable indicants of expertise. With some refinement of the observation and coding procedures, however, it may be possible to strengthen the correlations. Furthermore, although no variable by itself may be strong enough to identify level of expertise, some combination of the variables may yield a reliable prediction.

The CODE II experiment streamlined the methodology used in the CODE I experiment. The reliability analyses from the CODE II data suggest that the methodology can be further streamlined in several ways. First, the high inter-judge reliability indicates that the use of even a single judge would yield acceptably reliable expertise levels. Second, because the judges' ratings based upon the written products generated by each subject were highly correlated with their ratings based upon assessing the subject's decisionmaking process through viewing the videotape, the time-consuming videotape viewing could be eliminated and expertise level based solely upon rating the written products. (Note that the judges perceived the tapes as more useful than the written materials, however.) Even if some evaluation of process were deemed worthwhile, it is clear from the strong inter-correlations among the three process measures in CODE I and the two in CODE II that any one of them would be sufficient.

Considerable insight into the subject's decisionmaking expertise is available through the questions he asks. Instead of the free format we used in these experiments, one could give the subject a large set of questions from which he would choose a defined maximum to ask. This set of questions would have to be sufficiently rich that the included critical questions would not stand out. We could then assess the subject's expertise based on the questions he selected, using the judges' ratings of the criticality of the questions in the list to weight questions by degree of importance.

Issues for Further Research

There is a variety of potential follow-up activities to the CODE experiments, which we briefly discuss below:

RPD behavior. One aspect of our theory that was not significantly supported by the CODE experiments was the notion that the expert maintains an extensive store of specific experiences in memory, retrieves experiences relevant to a new situation quickly, and uses those experiences to generate an initial schema and plan — behavior termed recog-nition-primed decisionmaking by Klein following his observations of it. Although we observed that higher experts rapidly generated an initial COA that contained more detail and more contingencies, we were unable to tie this COA-generation explicitly to specific experiences in memory. More research is needed on how the expert stores his experi-ences and retrieves relevant aspects of that experience when confronted with a new situation.

The MCD expert and uncertainty. We found no direct evidence in the CODE ex-periments for a relationship between expertise and uncertainty, and therefore we were unable to clarify that aspect of our theory. We see a refinement of our theoretical frame-work, in which we view the expert as well calibrated relative to the actual uncertainty, but able to take decisive action as long as the uncertainty is acceptable, i.e., will not com-promise his mission. Because uncertainty is perhaps the most significant factor that makes the military commander's job difficult, this area is clearly worthy of further exploration. One approach would be to directly control the uncertainty in multiple situations (relying on one or more super experts for verification) and more directly examine issues such as perceived uncertainty, perceived difficulty, and confidence.

Secondary predictive measure development. A potential extension of the current framework is the refinement and expansion, through several experiments, of our current set of secondary measures into a set that could reliably be used, in place of the super-expert assessments, to define a subject's expertise level. While these secondary measures could be scored by military nonexperts, they would require the nonexperts to view videotapes, thus taking longer than a super-expert reviewing written material. Whether this would be an economical tradeoff, or whether such a set of secondary measures is feasible, remains to be investigated.

Additional analyses of the CODE experiment data. We have already noted that the data indicate it is possible to obtain a reliable evaluation of expertise with fewer judges and/or fewer ratings by the judges. The existing data could be reanalyzed to see whether the findings showing empirical support for our theoretical framework would be affected by using fewer judges or a subset of the ratings. For example, we could reanalyze the data basing the expertise ratings only on:

- an individual judge's ratings
- the judges' evaluations of the written materials[10]
- the judges' evaluations of the process measure assessing the subjects' initial reaction to the situation.

The results of such analyses could provide an indication of whether an experiment testing an even more streamlined data collection procedure would be fruitful.

It would also be possible to conduct a more systematic analysis of the judges' comments about their ratings. We have already done a qualitative analysis of the judges' comments for the subjects in the high- and low-expertise groups. We could develop a categorical coding system for

[10] An analysis of the relationship between the subjects' expertise scores derived only from the written materials and the nonexpert raters' measures has already been conducted and reported on. We could also do the comparison of the secondary measures and the judges' comments for the high- and low- expertise groups, where the identification of the members of the two groups was based only on the subjects' scores on the written materials.

the judges' comments that would make it possible to do a more quantitatively based analysis that could, for example, reveal whether the quantity of shortcomings noted is related to expertise. Systematically coding the judges' comments would also allow us to analyze the extent to which the judges used the same or different pieces of evidence to support their expertise ratings. We know that they generally concur in their ratings of the subjects, but at this point we have no quantitative measure of the degree to which they concur in the reasons for their ratings. Additionally, this could be investigated in the context of the few instances in which the judges were in significant disagreement about a subject.

It would also be possible to conduct a multivariable analysis of the secondary measures to see whether some combination of observable measures could reliably predict level of expertise. We could also try to derive additional nonexpert ratings by developing coding schemes for the subjects' responses to the open-ended questions.

Theory development. The CODE experiments have provided solid support for some aspects of our theory. Subsequent investigations would allow us to determine whether the hypotheses we could not verify in the CODE experiments can be empirically supported with supplementary analyses, a refined experiment procedure, and/or a different subpopulation.

An important issue for MCD-expertise development is the relative importance of an individual's innate ability versus his training and experience. To what extent must the expert commander be selected, rather than created through training? Although the evidence from the CODE experiments is sketchy on this question, it does provide some support for the importance of experience in MCD expertise: a large proportion of the more-expert subjects reported experience serving as S-3 or G-3 operations or planning officers, while few of the less-expert subjects reported such experience. Also, the critical comments made by the judges stressed factors that seem addressable through training, such as the need to coordinate maneuvers and fires. Although the judges sometimes mentioned seemingly innate qualities such as intuition or "warfighting instinct," the vast majority of their comments dealt with the need for a detailed, complete plan that made good use of available resources and provided the required guidance to insure coordination of those resources over time and space. Expertise theory, and our theoretical framework for MCD expertise, provide little guidance in this area. Future theory work should extend our framework to address this issue.

The data collected in the CODE experiments could be used to provide a preliminary test of an extended theory. The majority of subjects in the sample are of the same rank (major) and have roughly the same number of years of military experience, but differ in their experience. A fine-grained analysis of the data from this subsample, guided by an extended theory, could shed light on the contribution of innate ability and the relative contributions of particular kinds of experiences to MCD expertise.

Recommendations for Potential Applications

The CODE experiments answer two questions that have profound implications for the assessment and training of MCD expertise. First, can MCD expertise be measured? Second, can MCD skills and expertise be reliably elicited in a simple environment (i.e., written materials and maps) such as that used in CODE? Our results indicate that the answer to both questions is "yes."

A method of measuring MCD expertise is needed in order to identify the components of that expertise that are most often missing among Army officers in order to determine the kinds of experience or training that are most needed. MCD expertise measures are also needed to

determine what kinds of training and/or experience are most helpful in increasing expertise, and to assess the value of alternative training methods and programs. The CODE results establish that battle command decisionmaking expertise can be reliably measured, and that this measurement can pinpoint shortcomings in expertise. Many of the shortcomings identified were not isolated instances but were common among the officers who participated in the study.

There is also a pressing need for less-expensive, more widely available portable training methods that can supplement large-scale exercises and simulations. The CODE results establish that simple materials such as written descriptions and maps do provide a sufficiently rich environment for assessing expertise and for diagnosing the weaknesses of individual decisionmakers in order to provide feedback. The success of CODE in measuring MCD expertise and identifying shortcomings in MCD skills in an inexpensive and easily portable environment suggests a number of possible applications, discussed below.

Development of Command Skills for Army Officers

One application of the CODE methodology, materials, and results is to develop case-based training to improve MCD skills. A set of training situations, similar to those used in CODE, could be developed. As in CODE, each situation would have an associated list of potentially available information that could be gathered through questions. While there would be no single "right" answer for a situation, a number of acceptable plans would be developed, along with a list of the most-important questions that should be asked, the most important issues to be considered, and the key elements that should be present in the final course of action. As in CODE, officers would study the situation, gather information through questions, and develop a plan. They would then assess their plan by comparing the issues they considered and the plans they developed with those of expert commanders. Evaluation could be through self-assessment or could be provided by a trainer.

This scenario-driven, case-based training could be provided in the form of paper-based training materials to be incorporated into an existing training curriculum. Training packages could be used, for example, in command and control and decisionmaking courses at CGSC. Such packages would include instructor training in how to use the package as well as individual officer-training materials.

Another possibility is the development of a portable PC-based individual training tool that could be used in preparation for BCTP, CPX, or NTC exercises. This tool would present tactical situations electronically using maps, graphics, descriptions, etc. Officers would then develop a plan for responding to the situation. Once the plans were developed, the tool would provide self-assessment materials as described above, discussing the major issues that should have been considered and the major elements that should have been included in the plan. The PC-based tool could be used with or without the presence of an on-site instructor.

Decisionmaking Performance: Evaluation and Feedback

The CODE methodology could be applied for diagnostic assessment of the MCD expertise of individual commanders, providing feedback on the individual's strengths and weaknesses in assessing a tactical situation and developing a plan. It could also be applied to a group of individuals in order to assess the effectiveness of training programs. A pre-training assessment would identify the most common shortcomings in the group, indicating the areas where training should be focused. A post-training assessment would identify training shortfalls, identifying the areas where more training or experience is required, or areas where the training could be improved.

Our judges felt that by conducting a systematic evaluation of a range of subjects using the same stimulus materials, the CODE process has already uncovered some important shortcomings in officer training, as indicated by the following quotes:

> "Overall, subjects are weak in: use of fires to support maneuver, seeing the battlefield, realizing they are part of a larger force and therefore a larger fight, maintaining focus on mission, and articulating an intent that results in a 'shared mission' for the force."

> "I have already talked... about the difficulty many of [the subjects] have in putting their conceptual thoughts in writing. We will continue to pursue that issue. Another issue, perhaps the most significant one, is the difficulty that so many of the subjects encountered in understanding the intent of the higher commander. In light of the effort the Army has placed on this subject it's surprising that we are not doing better."

Human Performance Assessment in Distributed Battle Settings

The CODE methodology provides a method for assessing performance in distributed wargaming simulations such as DIS. CODE methods could be used to assess the strengths and weaknesses of officers coming into a simulation exercise and again after their participation in the exercise to evaluate the effectiveness of the simulation in increasing MCD expertise and to identify skill areas that are still in need of improvement.

REFERENCES

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper and Row.

Deckert, J. C., Entin, E. B., Entin, E. E., MacMillan, J., & Serfaty, D. (1992). Military command decisionmaking expertise--Annual interim report (Technical Report No. 568). Burlington, MA: ALPHATECH.

Entin, E. E., Needalman, A., Mikaelian, D., & Tenney R. R. (1988). Experiment II report: The effects of option planning and battle workload on command and control (Technical Report No. 388-1). Burlington, MA: ALPHATECH.

Johnson-Laird, P. N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Kahan, J. P., Worley, D. P., & Stasz, C. (1989). Understanding commander's information needs (Technical Report No. R-3761-A). Santa Monica, CA: RAND.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95, 163-182.

Klein, G. (1988). Naturalistic models of C^3 decisionmaking. In S. Johnson and A. Levis (Eds.), Science of command and control: Coping with uncertainty. Fairfax, VA: AFCEA International Press.

Nunnaly, J. (1967). Psychometric theory. New York: McGraw Hill.

Riesbeck, C. K., & Schank, R. C. (1989). Inside case-based reasoning. Hillsdale, NJ: Erlbaum.

Serfaty, D., Deckert, J. C., Entin, E. B., Entin, E. E., & MacMillan, J. (1993). Developing command decision-making expertise: Workshop report (Technical Report No. 581). Burlington, MA: ALPHATECH.

Serfaty, D., MacMillan, J., & Deckert, J. C. (1991). Toward a theory of tactical decisionmaking expertise (Technical Report No. 496-1). Burlington, MA: ALPHATECH.

Serfaty, D., & Michel, R. R. (1990). Toward a theory of tactical decisionmaking expertise. Proceedings 1990 Symposium on Command and Control Research, Monterey, CA, 257-269.

GLOSSARY

AAR	after-action review
AD	armored division
AMSP	Advanced Military Studies Program
average expertise rating	for each subject for each judge, the average of the component scores over the tactical situations
BCTP	Battle Command Training Program
bde	brigade
BOS	battlefield operating system, of which there are seven: command, control, and communications; intelligence; maneuver; air defense; fire support; mobility/ countermobility/survivability; and combat service support
CBS	Corps Battle Simulation
CG	commanding general
CGSC	Command and General Staff College
COA	course of action
CODE	COmmand Decisionmaking Expertise
component measures	judges' ratings of the subjects' written concept, written messages, initial reaction to the situation, decisionmaking process, and reaction to the new information
CP	command post
CPX	command post exercise
expertise assessment	for each subject, the overall expertise rating from a judge
expertise level	for each subject, the average of the overall expertise assessments over the judges
JESS	Joint Exercise Support System
JSTARS	Joint Surveillance and Target Attack Radar System
MCD	military command decisionmaking,

	incorporating both operational (corps and above) and tactical (division and below) elements
mean component score	for each subject, the average score of that component measure over the tactical situations and the judges
METT-T	mission, enemy, terrain (including weather), (own) troops, and time available
MLRS	Multiple Launch Rocket System
NTC	National Training Center, Fort Irwin, CA
O/C	observer/controller
OPFOR	opposing force
OPORD	operational orders
RPD	recognition-primed decisionmaking
SAMS	School of Advanced Military Studies
SITREP	situation report

APPENDIX A

Field Observations

Introduction

Experiments provide one way in which theories about MCD expertise can be empirically validated. Observations of realistic simulations and war games provide an alternative approach. Both simulation exercises and controlled experiments provide empirical data on decisionmaking expertise but they differ in their purpose and scope, and in the extent to which their results can be generalized beyond the events actually observed. Simulation exercises typically involve a number of interacting players dealing with a series of events in an organizational structure and environment that is as close as possible to an actual battlefield environment. While the outcomes of a simulation exercise may be measured by such factors as resource loss or enemy loss, there are so many uncontrolled variables in the exercise that it is extremely difficult for the untrained observer to attribute those results to any one of the many factors that can affect outcomes, such as individual abilities of the players, the training provided, the team structure used, the rules of engagement employed, or the intelligence support provided. It is difficult to judge whether the same results would be obtained if the same situation was replicated with other players. Nor is it possible to compare systematically the behavior of experts and nonexperts. However, simulation exercises provide a realistic environment for observing the theoretical components of MCD expertise and for formulating hypotheses than can be more rigorously tested in experiments. This section discusses the observation activities we conducted at a battle command seminar and a training exercise, in order to evaluate them as vehicles for extending and/or confirming our hypotheses on MCD expertise.

The Battle Command Training Program (BCTP) trains division and corps commanders and staffs against any threat worldwide using simulated engagements (Bartlett, 1989). A BCTP application consists of three phases: a battle command seminar, a command-post exercise (CPX), and a sustainment training package. The battle command seminar involves the commander and his primary subordinates in five days of workshops and decision exercises, for which they have prepared via a professional reading program. The emphasis of this phase is on team building and the development of war plans. Between two and six months after the seminar the CPX (Warfighter) occurs, consisting of five days of intensive battle based on the unit's war plans, simulated using the Corps Battle Simulation (CBS), a derivative of the Joint Exercise Support System (JESS). BCTP has developed a highly competent opposing force (OPFOR) group to provide a doctrinally realistic thinking and reactive enemy. After-action reviews (AARs) are conducted at appropriate times during the CPX. The sustainment training package is provided to the unit several months after CPX. This package contains three or four situations, tied to specific teaching points during the CPX, that the unit commander can use in his own seminar.

As part of our Phase II first-year activities we undertook to observe, evaluate, and assess MCD expertise characteristics during the first two stages of a BCTP exercise. To prepare for directed observations we enlisted the services of a consultant, an experienced commander, to support our development of a set of behaviors relevant to MCD that we might observe. The decision exercises during a unit's BCTP seminar seemed to offer an observation opportunity during which we could assess possible techniques for measuring behavioral components of expertise in an ongoing process. The Warfighter exercise would then provide an opportunity to test these techniques in a near-real-time (albeit simulated) combat situation.

Overall, we set out to achieve several objectives during our observation of the BCTP seminar and Warfighter. One was to gather information about the ways in which the components

of MCD expertise are expressed in a military exercise. A second was to test the preliminary observation instrument that we had devised for recording behaviors hypothesized to be related to MCD expertise. In particular, we wanted to ascertain which behaviors associated with MCD expertise could be recognized and recorded by military-nonexpert observers. A third was to use the exercises as a basis on which to generate and refine a procedure for eliciting and measuring MCD expertise in laboratory experiments.

BCTP Battle Command Seminar

We attended a BCTP battle command seminar for students enrolled in the School of Advanced Military Studies' (SAMS) Advanced Military Studies Program (AMSP) at Fort Leavenworth, KS. The team of observers for the project included two members of the ALPHATECH staff and a representative from the Army Research Institute Fort Leavenworth Field Unit.

The seminar was comprised of two interleaved activities: a battle planning exercise and workshops. The battle planning exercise was designed by the BCTP and carried out by the AMSP students under the guidance of a team of BCTP observer/controllers (O/Cs). The senior O/C for the BCTP was a retired general with extensive experience in MCD. The workshops, presented by senior military officers (most of whom were members of the BCTP staff), provided the students with information necessary or helpful for carrying out the exercise. The particular BCTP seminar we observed was different from the typical one in that the participants were SAMS students rather than the members of an existing unit who were already assigned to particular positions on the staff.

Battle Planning Exercise

The battle planning exercise was carried out by the students under the auspices of the BCTP O/Cs. The 52 AMSP students were divided into two divisions, a light and a heavy division. Within each division, each student was assigned a particular role (e.g., division commander, chief of staff, G2, G3, engineer).

The exercise was comprised of four subactivities: 1) mission analysis; 2) course of action (COA) analysis and comparison; 3) development of operational orders (OPORDs); and 4) development of situation reports (SITREPs). Each of the four subactivities was carried out in parallel by the two divisions. After each subactivity was completed, each division, in turn, presented a briefing to the division commander and to an audience comprised of the entire group of students, their SAMS instructors, the BCTP observer/controllers, and our two observers.

Observation of the battle planning exercise allowed us to see how tactical analysis and planning is carried out by a staff and how the results of this process are described and communicated by the members of a division staff to the division commander. Observation of the senior O/C's comments allowed us to see the extent to which the theoretical components of expertise that we have hypothesized are discussed, taught, and modeled by an expert commander.

Mission Analysis. Five days prior to the seminar, the students were given the OPORDs for the exercise. They were also given the corps commander's intent and corps staff brief. Between the time they received this material and the start of the seminar, the two divisions prepared a mission analysis. The presentation of the mission analysis briefings was the first student-led activity that we observed during the seminar.

The mission analysis briefing (and the other subactivity briefings as well) involved presentations from the viewpoint of the various members of the commander's staff. The mission analysis briefing included an analysis of the enemy (location, strength, plan of attack, vulnerabilities) and friendly goals and resources (tasks, forces available, restrictions, risks, and timeline). After the analysis from the various perspectives (e.g., personnel, logistics), the central presenter (in one group the Chief of Staff and in the other the G3) offered the mission statement recommended by the division staff. The division commander could change, modify, and/or approve the mission statement. Once the mission statement had been finalized, the division commander provided his guidance for the development of COAs, the next activity for the staff.

During the briefing and at its conclusion, the senior O/C offered a number of "teaching points." These were points he wanted the students to consider, or sometimes points he wanted to reinforce. Based on our conversations with the SAMS instructors and with some of the O/Cs (all of whom had already been through BCTP seminars), we concluded that they too learned from the senior O/C's comments and questions.

The issues that the senior O/C touched on during his discussion of the mission analysis offer validation for our hypotheses about MCD expertise. Among the topics he discussed and the components of our theoretical framework that they substantiate are the following:

- Terrain: "Is there decisive terrain (terrain you have to have to complete your mission)?" This question offers support for the hypothesis that experts generate a mental model of the situation and that they fill the unknowns in the model through information gathering.

- Intel: "Find out what is real and what is templated". This point supports the hypothesis that experts seek disconfirmation of information (in this case actual versus expected information).

- Looking from different points of view: "Learn to look from the operational level as opposed to the tactical level." This admonition supports the hypothesis that experts build and use a richer mental model of the situation.

- Coordination between units: "You are part of a mosaic. You aren't a free agent." The senior O/C noted that one group's goals may be in competition with another group's goals and he discussed the inherent competition between a unit's need to "look for the edge" and the greater good. These comments support the hypothesis that experts build teams.

- Show stoppers: After both divisions had finished their briefings, the senior observer asked the students "what are the show stoppers for both divisions?" This question offers support for the hypotheses that experts look from the enemy's point of view and that they visualize (potential) outcomes.

At the end of this section of the seminar the senior O/C praised one division commander on the presentation of his guidance to the staff. We found that we, as military-nonexpert observers, were not able to discern that one presentation was better than the other. When we asked the AMSP instructors why this presentation was better, they told us that the other commander's presentation was much too detailed. We concluded from this experience that as military-nonexpert observers we could not evaluate the quality of a military presentation.

COA development and comparison. At the conclusion of the mission briefings, the division staffs were charged with the development and comparison of alternative COAs. This procedure involves specifying one or more COAs, "wargaming" each COA, and then comparing the COAs

on a number of criteria, with the overall goal of recommending a COA to the division commander. We observed this activity in an attempt to see if we could identify and record behaviors hypothesized to be associated with MCD expertise.

As part of our preparation for the seminar observation activity, we developed a preliminary observer form that we hoped to use to record instances of behaviors that our military consultant had helped us identify as aspects of MCD. The COA development activity was the first opportunity we, as observers, had to watch the staff as they worked interactively, in order to see if we could observe evidence of and record occurrences of behaviors we had enumerated in the preliminary observer form. In particular we were looking for evidence of aspects of information seeking (for example, about the enemy, own troops and supplies, terrain, and weather) and information giving (for example, communicating mission or intent, evaluating situation or plans, or differentiating templated from known information). Trying to use our preliminary observer form allowed us to assess whether such behaviors were observable and whether they could be recognized and captured in real time by military-nonexpert observers.

We found that it was occasionally possible to identify instances of the behaviors that we had enumerated in our preliminary observer form, but that in some cases it wasn't clear what category a behavior fell into (for example, was the commander questioning the veracity of a subordinate's statement or asking for clarification of the statement). We also found that the extensive use of acronyms and synonyms made it difficult for military-nonexpert observers to understand and categorize statements, especially in a fast-paced discussion. (Note that the use of nonstandard terminology is a common deficiency revealed in BCTP applications.) The problem of clearly defining the meaning and boundaries of behavioral categories can only be overcome by extensive training and calibration among observers.

Perhaps more important, though, there was no way for us to evaluate the <u>appropriateness</u> or the <u>quality</u> of a behavior, and this is not necessarily a problem that observer training would solve. For example, if the commander evaluates the worst-case outcome for some action, we do not have the ability to judge the quality of his evaluation, nor can we judge whether this was an appropriate time for the commander to state his evaluation. If a commander prods for additional information, we cannot assess whether he is seeking relevant information or whether he is overly concerned with details.

There is also a question of which level of expertise we are studying. Note that the COA-development process is not usually run or "chaired" by the division commander (the role that we had targeted as the focus of the data collection). In the two divisions we observed, one COA-development process was explicitly run by the G3 and the other was run de facto by the two most vocal staff members. During a planning period such as this, the division commander, himself, might be out "visiting the troops" (for example, talking to brigade commanders). If the focus of the observation process was the division commander, then the observer would have to have the ability (including both permission and means) to follow the division commander as he went about his activities.

COA development and evaluation occupied all of the ensuing afternoon and evening. Another problem that we encountered as observers is that after about three hours, the division organization degenerated and the group tended to splinter. This splintering of the group impeded our attempt at recording group interaction.

The following morning each division presented its COA development and evaluation briefing to the division commander and the assembled group of students and observers. In the presentation, each COA was described and its advantages and disadvantages were evaluated from the point of view of the various subunits (intel, maneuvers, fire support, logistics). After all the COAs had been presented, they were compared in terms of such factors as simplicity, mass, combat power concentration, risk, flexibility, and end state, and one was recommended to the

division commander. Following the recommendation of a COA and its approval and/or modification by the division commander, the commander gave his guidance for the development of the OPORDs.

Again, during the presentation of the COAs and after both divisions had concluded their presentations, the senior O/C offered teaching points. One of the major points of discussion revolved around the idea of a "stop," a tactical pause in which a unit rests and regroups. The senior O/C made the point that once you pause the troops, it is very hard to get them going again (both physically and psychologically). But on the other side, there is doctrine (and experience) which says that the soldiers cannot keep going indefinitely. The senior O/C's evaluation of both the pros and cons of a stop offered validation for our hypothesis that experts take into account both the physical and psychological conditions of their own troops. The positive and negative effects of pauses were illustrated through the use of experiences (war stories), offering validation for our hypothesis that experts maintain an extensive store of experiences in memory.

Another point that the senior O/C stressed was the necessity of visualizing two levels down (in this case to the brigade and battalion levels). He emphasized that although you don't want to tell the commanders at those levels how to fight, you need to visualize what they will do, so that you can see what resources they will need (as everywhere else, the various units compete for resources). This illustrates the importance of an expert's richer mental model of the situation, in this case the expert's ability to visualize the battlefield and enumerate the resources that will be required by a subordinate commander in order to achieve a desired outcome.

<u>Development of OPORDs and SITREPs</u>. The next afternoon and evening were devoted to the development of the OPORDs for the chosen COA. From what we could observe there were no organized staff meetings. Not all the seminar participants were present at one time, and those who were present seemed to be meeting in small, amorphous groups. As a result there was no coherent opportunity to observe the divisions as they developed the operational orders.

The briefings of the OPORDs occurred the following morning. The briefings were organized and presented by phase of the operation, and included task orders for each phase. Here again there was discussion of the potential negative effects of operational and tactical pauses. "When you stop a unit it is pure hell to get it started again." There was also discussion on synchronization difficulties and about key decision points. The focus on key decision points offers evidence that development of expertise involves learning to ask the "right" questions and perform the "right" analysis.

After the operational orders were discussed, the CBS was run, using the operational orders produced by the two divisions and the OPFOR plan. Very early on the following morning, based on the results of the computer simulation, the division leaders received an update from the corps staff. This update indicated what had occurred in the simulated battle, and the situation of their divisions. After the update, each of the divisions was given one hour to prepare a SITREP for its division commander. The preparation of the SITREP was done by a small group of the staff and was not opened to observers. The staff presented the SITREP to the division commander (and the audience), after which the division commander gave his guidance.

The discussion that followed the SITREP and commander's guidance, shaped by the senior O/C's comments, centered around what the enemy thinks is the main thrust of the friendly attack, and what friendly now thinks about the enemy. The nature of this discussion supported the hypothesis that the expert looks at the situation from the enemy's point of view.

<u>After Action Review</u>

The last part of the seminar was an AAR, conducted by a member of the BCTP staff. The AAR gave the students an opportunity to "revisit what was accomplished with a focus on the decisionmaking process." The purpose of the AAR was not to evaluate the plans generated by the students, because you "cannot evaluate a tactical plan without playing it out." Much of the discussion focused on wargaming, "the most difficult and time consuming process." In general, the discussion was fairly detailed and, although it made us cognizant of the layers of detail embedded in expertise, it was not directly related to our observation activity.

Workshops

Nine workshops, each between one and two hours long, were presented during the course of the seminar. The workshops educated us on a number of aspects of battle planning. The workshop on preparing for a Warfighter exercise was helpful in preparing us for our upcoming observation of a Warfighter in that it helped to shape our expectations of what would occur there. The workshop on leadership was interesting in comparing requisite qualifications for and characteristics of military leadership to leadership in other, nonmilitary domains.

Conclusions about the Seminar-Observation Process

We found that even in the semi-structured environment created by the seminar, it was difficult to recognize and categorize a predefined set of behaviors hypothesized to be relevant to MCD expertise. Observing this seminar did provide us with background for observing an actual Warfighter, but it did not allow us to refine our preliminary observation form such that it contained a set of measures that we were confident we could observe during a Warfighter. In fact, our experience reinforced our estimation of how difficult it is to make observations in a fluid situation such as the Warfighter.

The purpose of the battle command seminar is not to evaluate the participants, but rather to teach them about battle command and to prepare them for a Warfighter. As a result, there is no evaluation in terms of how good their plan was, or how well they had played their roles. Thus there is no way to correlate a set of behavioral observations with greater or lesser expertise — we don't know who are the experts.

BCTP Warfighter

A Warfighter exercise provides a realistic environment that cannot be captured in an experiment setting, and offers the opportunity to observe the real-time interaction of commanders with their subordinate commanders, with higher level authorities, and with collateral units. We hoped to observe manifestations of MCD expertise by focusing on the commander's decision process and his interactions with members of his staff.

The BCTP Warfighter exercise we observed involved an infantry division. The "war" commenced on a Sunday evening, and was terminated early on Thursday. The BCTP team and the CBS were housed in the simulation center. The battalion leaders, who entered their orders into the CBS, were also located here. Separated from the simulation center and from each other by several kilometers were the various command posts (CPs) of the division and the brigade headquarters.

Simulation Center

The lead representative at the exercise from the contractor that supports CBS conducted us on a tour of the simulation center. He explained to us that an exercise is organized to play three echelons of command, in this case division, brigade, and battalion. As noted above, division and brigade headquarters are located in the field, but battalion headquarters are simulated in the simulation center. The battalion leaders (with the help of technicians) enter into the simulation all orders and information that have to go down the chain. Leaders at the battalion level "see the war" — e.g., their computer screens actually include maps with icons placed on them. Brigade- and division-level players get reports.

Each battalion, brigade, and division unit is supervised by a BCTP O/C who serves a dual function: as an umpire (making sure doctrine is followed) and as a mentor. Most O/Cs are either active or retired members of the military. The senior O/C for the exercise serves as a mentor to the division commander. It was emphasized that nothing is written down — everything is done at the personal one-on-one level. For example, each day the senior O/C takes "a walk in the woods" with the division commander. In addition, he spends about a half hour each day with brigade and battalion commanders, giving them what might be called tutorials (e.g., "did you consider this", "what about ..."). The retired general who was the senior O/C at this Warfighter was aware of and supportive of our project, and had in fact encouraged us to observe a Warfighter.

The OPFOR, headed by a member of the BCTP staff, involves over 100 people (located at Ft. Leavenworth). The corps commander (of the training division) can "shape" the exercise by manipulating the information available to OPFOR and/or the training division. The information from corps to the division is simulated by the BCTP staff, in accordance with the corps commander's intent, which was spelled out some months prior to the Warfighter when the division began the planning phase of the battle planning seminar.

In the course of our tour of the simulation center, our host, who was a retired military officer, noted that a commander's orders to a subordinate commander are clearest when they are given "on site." with the "lay of the land" in front of them. He gave an example of a division commander's telling a brigade commander to "hold that hill." The brigade commander has to interpret what this means (for example to put men on the hill or to block the enemy from taking the hill). If the two commanders are on site, the brigade commander may then respond by explaining what he will do (e.g., "ok, I will send my men along that road"), whereupon the division commander says, "no, you don't understand what I mean, just ..."). This conversation emphasized for us the importance of visualizing the battlefield, and the importance of aids, such as a view of the actual battlefield or maps, in that process.

Observations at the Command Posts

By visiting the various command posts we hoped to be able to observe behavioral components of command decisionmaking during an ongoing exercise. This subsection discusses our experiences at the command posts.

D-Main. The first command post we visited was D-main, the main division CP, which was comprised of five functional units (current ops, intel, plans, fire support, and air support) and the commander's briefing room, located in six adjoining trailers. The trailers were small, with little floor space in them, and it was not easy to fit in an extra body without potentially being in someone's way. As observers, we were not permitted to enter the division commander's briefing room (clearly a likely place to observe high-level decisionmaking). Because of the crowded conditions, it was suggested that only one of us should be in a trailer at any particular time.

In our first visit to D-Main, one of us observed in current ops and the other in plans. We spent more than half a day in these positions, with occasional migrations to the other trailers at D-Main. With several intercom nets being on all the time, the current ops trailer was very noisy and

it was extremely difficult to hear what was being said, especially if the observer was more than one or two feet from the speaker. For any conversations on the telephone, one could only guess (or at best infer) who was speaking and what was said on the other end of the line. The planning unit was somewhat quieter (they had no intercom nets available), but one could only observe activity at a fairly low level. The individual soldiers in that unit were very cooperative and helpful, and would answer any questions we had, but they were not the high-level division leaders (many were not even officers).

Based on our experience at the SAMS seminar, we had revised our preliminary observer form. There were some behaviors (e.g., statements) that we could put into one of the categories on the form (e.g., requesting information about enemy location) but, as at the SAMS seminar, most of the statements were either ambiguous (in terms of the category into which they fell) or unintelligible. Furthermore, any decisionmaking we saw was at a very low level.

We revisited D-Main a second time with similar results. One of us was able to observe a portion of a planning briefing for the division and corps commanders, but without sufficient information about the situation and a knowledge of the participants' roles, it was not possible to make meaningful observations. Furthermore, we still did not have access to the commander's briefing room.

<u>D-Tac</u>. The division tactical CP, D-Tac, is normally located about 15 km from the front, and is primarily concerned with the immediate battle, as opposed to D-Main which is concerned with the battle 18-36 hours out, and D-Rear, which is concerned about security in the rear. During the time we were there, the members of the staff were working, but there was no observable decisionmaking behavior that could be recorded, and no high-level decisionmakers were present.

The chief O/C at that location has worked as an O/C for four years. He spoke to us about the kinds of errors that he has observed. One source of error he discussed is that the staff at D-Main can become too focused on the immediate battle, and therefore not be thinking about what will happen farther out (as in fact occurred in this exercise). Another source of error he noted is when commanders do not use accepted terminology, and try, instead, to invent their own terms (he speculated this is because they want to leave their own imprint). Although their own unit may understand them, outsiders will not, and this can lead to incorrect intent.

We asked the chief O/C what makes a good commander. Among other qualities, he said good commanders seem to have a sense of something's being "not right," and that they know when to step in. This speculation is consistent with the hypothesis that the expert has a <u>more detailed mental model</u>. Because they are attuned to the functional relationships, they may see small perturbations which others wouldn't notice, and which they translate to "something's not right." He also felt that potential experts can be identified early on by older soldiers, but not necessarily by their own peers.

<u>1st Brigade TOC</u>. After visiting D-Tac, the senior O/C led us to the 1st Bde Tactical Operations Center (TOC), where, again, the senior O/C talked with the soldiers and we were put in the charge of one of the O/Cs at that location. The O/C asked the captain in charge to talk us through the functions of the TOC (which replicate those of the division on a smaller scale). While we were being shown around the TOC, the brigade commander walked in. He was willing, and had the time available, to talk with us for a short while. He explained his view of how the division had run into difficulty early in the exercise. That is, the division pushed the enemy back, and never thought of the consequences of what would happen or what it would do once it took the high ground. He said the division was up there without its own artillery support, facing an enemy with greater range and more artillery. What occurred is that the division got hammered by the artillery. What the division should have done, he said, once it stopped the enemy there, is to have

pulled back in an orderly manner, and forced the enemy to meet it. In terms of the hypotheses about the nature of expertise, one might suggest that the disaster was caused by not looking from the enemy's point of view.

Rock Drill

A rock drill allows the commander and his high-level staff to walk through their plans. A rectangular piece of ground (about 10 to 15 feet on a side) is cordoned off and on it the phase lines and likely enemy attack corridors are laid out with colored rope. The commander talks and walks through the plan with his subordinates, making sure everyone understands the commander's objectives. The subordinate commanders stand at or move to the positions their units will attempt to occupy. (It is called a rock drill because they once used rocks to represent the units — now they use the live representations of the unit instead.) Attention is given to synchronization and who will control what resources. The subordinate commanders may be asked to verbalize their mission or to say what problems they think they will encounter. They go through all phases of the plan, including the attack and fallback.

We were permitted to observe the second rock drill, held after initial planning was completed. There were quite a few people at the rock drill, and it was not always easy to hear what the participants were saying. Because most of it was conducted after dark, with the only light being provided by the headlights of three or four vehicles, it was also very difficult to see the participants and the relative location of the terrain on which they were standing.

After-Action Review

After the war was halted we viewed the AAR on a monitor in the "overflow room." Among the problem areas discussed at the AAR were synchronization, transition from one phase of battle to another (e.g., offense to defense, or vice versa), and knowing who is in charge. In the course of the discussion, the division commander noted some things he had done wrong such as mistaking one kind of action for another. He also noted potential improvements, such as the need to talk about the engineering plan with the same degree of detail that they talk about fire support or maneuvers. He mentioned the need to be clear about who is in charge of executing a plan and the need for "visualization of the transition."

There was some discussion of who is responsible for picking out the commander's intent. It was suggested this is the responsibility of the subordinate commanders who need "to pull the intent out of the commander."

During the AAR the corps commander asked whether the division commander saw the primary purpose of the rock drill as synchronization or as an opportunity for the commander to talk through and rehearse his plan. The division commander said that it is a synchronization drill first, and "then we try to retrofit it a second time through." Another key question, the corps commander said, is who is playing the part of the enemy. Another person in the audience quoted the adage "a picture is worth a thousand words" and said that the rock drill "kind of sinks it in mentally." The chief engineer commented that during the rock drill the division commander's "intent became much more clear to me and we made some changes to the plan." Again, these comments reinforced the importance of visualization.

Difficulties Encountered in Observing the Warfighter

Access. Our initial contact at the Warfighter was not a member of the BCTP staff, and his attempts to gain our entrance to various facilities was fruitless. It was only through the senior O/C's influence that we were given some freedom of movement, and even he did not give us assured (legitimate) access to the high-level decisionmakers. In order to have full freedom of

movement, our presence (and our purpose) would have had to be made known to the high-level decisionmakers ahead of time, and they would have had to acknowledge and concur to our presence. In other words, someone with authority and stature would have had to have talked to them, allayed their fears, and secured their endorsements. The high-level division officers involved in the exercise we observed seemed suspicious of the outside observers, and they looked at us as intruders who got in the way and muddled the process.

The observation process. It became clear that it is difficult for military-nonexpert observers to learn much about MCD expertise at an exercise. Although observing an exercise may be helpful for generating hypotheses to investigate or for giving us a sense of the way in which a scenario unfolds, a laboratory environment permits systematic investigation of specific issues or questions. In order to make meaningful and reliable observations about decisionmaking during a Warfighter (or similar exercise), we conclude that it would be necessary to have an observer positioned close at the commander's side, taking notes, and to tape everything that a commander said or heard. These steps are required to fill the information gaps we encountered due to ambient noise and numerous rapid-fire telephone conversations.

Interpreting the data. Even if all these information-related problems were resolved, a military nonexpert could only record, but not categorize or evaluate, the commander's actions. Without the ability to place them in a larger context, there is no way to evaluate these actions. Placing the actions we saw in a larger context would require being knowledgeable about the intentions and plans of both OPFOR and the training unit. Being able to interpret the actions would require having continual access to a military-knowledgeable colleague who could place the actions being observed into the larger context, help the observer interpret military jargon, and, hopefully, evaluate the quality of the decisions being made.

While we had hoped that the senior O/C could support us as such a colleague during our observations, it was clear that his job at a Warfighter leaves virtually no time for such non-BCTP activities.

Conclusions and Recommendations

While the primary purpose of a Warfighter exercise is to train commanders and their staffs, there is no reason why an exercise cannot also be used to collect data about MCD expertise. It should be feasible to meet both training and research objectives if the following recommendations can be met:

- The group conducting the exercise must agree to allow the observers to attend the exercise and must provide them with ongoing access to high-level information about the goals and the progress of the exercise. In order to anticipate and be prepared for potentially significant events, observers need to be aware of the OPFOR's plan.

- The participants in the exercise must agree to allow observation of all facets of the exercise. Those decisionmakers who are being observed must agree to permit the observers to follow them wherever they go and to stay close enough to hear everything they say.

- The observers must have access to all the information available to the decisionmaker. For example, they must be able to hear both sides of a telephone conversation, and be able to read all memos sent by and to the decisionmaker.

- The observers should have access to a military-knowledgeable colleague who can interpret language and events that the military-nonexpert observer cannot understand, especially in a fast-paced environment.

We were encouraged that many of our observations supported our hypotheses on MCD expertise, as we have noted throughout the section. Because of the aforementioned difficulties with currently observing a BCTP seminar or Warfighter, however, we feel that the remaining resources on this project are better invested in activities other than exercise observation.

APPENDIX B

EXPERIMENT MATERIALS

Judges Rating Forms

1. COA Rating Form (Intent and Messages)
2. Process Rating Form (Initial Reaction, Decision Process, Response to New Information)
3. Overall Rating Form

Subject ID _____ Situation _____ Date _____ Judge _____

COA Rating Form

1. Examine the intent statement carefully. Using the scale below, rate the tactical decisionmaking expertise exhibited by the intent statement (you can mark anywhere on the scale).

1	2	3	4	5	6	7
novice						expert

Please explain the positive or negative factors that influenced your rating:________________

2. Examine the COA embodied in the message(s) the subject has written. On the scale below, rate the tactical decisionmaking expertise exhibited by the messages.

1	2	3	4	5	6	7
novice						expert

Please explain the positive or negative factors that influenced your rating:________________

Subject ID _____ Situation _____ Judge _____

PROCESS Rating Form

Rating 1: Initial Reaction

Listen to the subject's initial reaction to the situation. On the scale below rate the degree of tactical decisionmaking expertise exhibited by the initial reaction (you can mark anywhere on the scale).

1	2	3	4	5	6	7
novice						expert

Comments:

____

____

____

____

Rating 2: Decision Process

Listen to the subject's decision process, including his questions to the experimenter, his verbal summary of his COA and his responses to the experimenter's questions about the tactical situation. On the scale below rate the degree of tactical decisionmaking expertise exhibited by the subject's decisionmaking process.

1	2	3	4	5	6	7
novice						expert

Please list at least two positive factors and at least two negative factors that influenced your rating:

Positive Factors	Negative Factors
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________

Additional comments about your rationale ______________________________

____

____

____

Subject ID _____ Situation _____ Judge _____

Rating 3: Response to New Information

Review the subject's response to whether the New Information about the tactical situation would cause him to modify his COA and his response to the experimenter's questions about the New Information. On the scale below rate the tactical decisionmaking expertise exhibited by the subject's responses.

1	2	3	4	5	6	7
novice						expert

Comments about your rationale ____

____

____

____

____

Subject ID ____ Date _____ Judge _____

Overall Rating Form

Please rate the subject's overall level of tactical decisionmaking expertise.

1	2	3	4	5	6	7
novice						expert

What are the two most important factors influencing your rating?

1.____

____

____

____

2.____

____

____

____

Comments:

____

____

____

____

____

____

____

____

____

____

____

____

Rater's Evaluation Form

1. CODE I Experiment
2. CODE II Experiment

Subject ID ___ Situation ____ Rater ___

RATER'S EVALUATION FORM

I. THE INITIAL COA

1. Sum up the main points of the subject's reaction to the "what are you thinking" question?

2. Counter Number for first evidence of COA ___

   Counter Number for Experimenter's Prod for COA (Thoughts on what you will do) ___

3. Did the subject provide a COA? ___ Yes ___ No
   If no, omit rest of Question 3

   3a. Did he volunteer the COA? ___ Yes ___ No

   3b. Detail of COA no detail |_____|_____|_____|_____|_____|_____| great detail

   3c. Is the COA linear or is there some evidence of contingencies?
   ____ linear ____ contingencies
   Evidence for contingencies

   ____

4. Did the subject explicitly flag anything as critical? ___ Yes ___ No
   If yes, what is flagged? ____

5. Approximately what proportion of the time did the subject use any of the wall maps:

   as he studied the situation

   no time at all |_____|_____|_____|_____|_____|_____| all the time

   in explaining his initial COA?

   no time at all |_____|_____|_____|_____|_____|_____| all the time

6. Summarize the subject's initial COA.

Subject ID ___ Situation ____ Rater ___

II. THE QUESTION PERIOD

1. List the questions that the subject asked.

2. Describe when in the planning period questions were asked:

_____ Virtually all questions asked in the beginning
_____ Most questions at beginning with the rest occurring
    _____ evenly spaced over time period
    _____ mostly at the end of the time period
_____ Some questions at the beginning and some at the end
_____ Questions evenly spaced
_____ Other: Please describe distribution ______________________________________

Subject ID ___ Situation ____ Rater ___

III. SUMMARY OF COA

1. To what degree does the final COA match the initial COA

   no correspondence at all |____|_____|_____|_____|_____|_____| perfect match

2. To what extent was the COA modified by responses to questions asked?

   no modification at all |____|_____|_____|_____|_____|_____| highly modified

3. Is the COA linear or is there some evidence of contingencies?

   ____ linear ____ contingencies

   Evidence for contingencies

   ____

   ____

4. To what extent does the COA take account of time dependency and event sequencing?

   not at all |____|_____|_____|_____|_____|_____| to a great extent

   Comments:

5. Summarize the COA

Subject ID ___ Situation ____ Rater ___

IV. RESPONSES TO INTERVIEWER'S QUESTIONS

1. What was the most critical aspect of the situation?

   1a. Does the subject's response correspond to questions he asked during the Q and A period?
   ___ Yes ___ No ___ Uncertain (explain) ____________________

2. What was the most critical uncertainty, initially?

3. Number of other aspects of situation considered in formulating COA? (do not include most critical aspect in count) ___________

   List aspects considered (continue on back of page if necessary)

   1 ____
   2 ____
   3 ____
   4 ____

4. Did anything in this situation remind subject of a previous experience? (Note: Check yes only if subject refers to something specific) ____ Yes ____ No

   4a. If yes, summarize response

5. Information most influential to subject in reaching a COA. Provide a condensed list.

6. What did subject enumerate as a show stopper?

7. Summarize subject's alternative COA

8. Transcribe the subject's rationale for choosing the COA he did

Subject ID ___ Situation ____ Rater ___

V. RESPONSE TO NEW INFORMATION

1. Before listening to this part of the tape, please rate the extent to which the subject:

1a. Anticipated the new situation

not anticipated at all |_____|_____|_____|_____|_____| completely anticipated

1b. Planned for the new situation:

not planned for at all |_____|_____|_____|_____|_____| completely accounted for in plan

2. To what extent does the newly revised COA agree with the summary COA? That is, to what extent did the subject change his COA?

no change at all |____|_____|_____|_____|_____| complete revision

3. What was the most critical aspect of the new information?

The VII CORPS commander's mission is:

to penetrate and envelop the Iraqi forward defenses, to quickly close with and destroy the Republican Guard Forces Command, and to shut tight the trap on Iraqi forces.

His instruction to the 1st AD was:

On G-1 close to Iraqi-Saudi border. Attack 0400 G-day in zone. Orienting on objective Purple, then Objective Collins, destroy enemy forces in zone. Prepare to attack forward in zone.

4. To what extent did the subject voice the importance of not compromising the mission?

not at all |_____|_____|_____|_____|_____| to a great extent - overall

Subject ID ___ Situation ____ Rater ___

RATER'S EVALUATION FORM

I. THE INITIAL COA

1. Time/Counter Start _______

2. Time/Counter Number for first evidence of COA ___

   Time/Counter Number for Experimenter's Prod for COA (if needed) ___

3. Did the subject provide a COA? ___ Yes ___ No
   If no, omit rest of Question 3

   3a. Did he volunteer the COA? ___ Yes ___ No

   3b. Detail of COA no detail |_____|_____|_____|_____|_____|_____| great detail

   3c. Is the COA linear or is there some evidence of contingencies?
   ____ linear ____ contingencies
   Evidence for contingencies

   ____

4. Did the subject explicitly flag anything as critical? ___ Yes ___ No
   If yes, what is flagged? ____

5. Approximately what proportion of the time did the subject use any of the wall maps:

   as he studied the situation

   no time at all |_____|_____|_____|_____|_____|_____| all the time

6. Summarize the subject's initial COA.

Subject ID ___ Situation ____ Rater ___

II. THE QUESTION PERIOD

1. List the questions that the subject asked.

2. Describe when in the planning period questions were asked:

_____ Virtually all questions asked in the beginning
_____ Most questions at beginning with the rest occurring
    _____ evenly spaced over time period
    _____ mostly at the end of the time period
_____ Some questions at the beginning and some at the end
_____ Questions evenly spaced
_____ Other: Please describe distribution ________________________________

III. SUMMARY OF COA

1. To what degree does the final COA match the initial COA

   no correspondence at all |____|_____|_____|_____|_____|_____| perfect match

   Notes ____

2. To what extent was the COA modified by responses to questions asked?

   no modification at all |____|_____|_____|_____|_____|_____| highly modified

   Notes____

3. Is the COA linear or is there some evidence of contingencies?

   ____ linear ____ contingencies

   Evidence for contingencies

   ____

4. To what extent does the COA take account of time dependency and event sequencing?

   not at all |____|_____|_____|_____|_____|_____| to a great extent

   Notes____

5. Did anything in this situation remind subject of a previous experience? (Note: Check yes only if subject refers to something specific) ____ Yes ____ No

   5a. If yes, what (history, text, training) ____________________________________

6. How many show stoppers did the subject enumerate?_____

Notes:____

____

____

The VII CORPS commander's mission is:

to penetrate and envelop the Iraqi forward defenses, to quickly close with and destroy the Republican Guard Forces Command, and to shut tight the trap on Iraqi forces.

7. To what extent did the subject voice the importance of not compromising the mission?

   not at all |_____|_____|_____|_____|_____|_____| to a great extent - overall

Notes: ____

Subject Questionnaires

1. Tactical Situation Questionnaire
2. New Information Questionnaire
3. End of Experiment Questionnaire

Tactical Situation Questionnaire

Situation ___ Subject ID ___ Date _______

Please evaluate the tactical situation on the following scales. Put an X on each scale where it best reflects your opinion.

1. How complex was this tactical situation?

|_______|_______|_______|_______|_______|_______|_______|

not complex at all — extremely complex

2. Initially, how much uncertainty was there in the tactical situation?

|_______|_______|_______|_______|_______|_______|_______|

no uncertainty at all — extremely high uncertainty

3. How much uncertainty is there in the tactical situation now?

|_______|_______|_______|_______|_______|_______|_______|

no uncertainty at all — extremely high uncertainty

4. Of the information about the tactical situation you would have liked to have, what percentage were you able to obtain?

|_______|_______|_______|_______|_______|_______|_______|_______|

0 % — 25 % — 50 % — 75 % — 100 %

5. How confident are you that your COA can deal with the tactical problem posed in this situation?

|_______|_______|_______|_______|_______|_______|_______|

not confident at all — extremely confident

6. How difficult was it for you to reach a COA?

|_______|_______|_______|_______|_______|_______|_______|

not difficult at all — extremely difficult

7. How adequate was the time allocated to develop your COA and write your intent and messages?

|_______|_______|_______|_______|_______|_______|_______|

much shorter than needed — much longer than needed

New Information Questionnaire

Situation ___ Subject ID ___ Date _____

Please evaluate the new information according to the following scales. Put an X on the scale where it best reflects your opinion.

1. How complex was the situation created by the new information?

|_________|_________|_________|_________|_________|_________|_________|

not complex at all	extremely complex

2. How difficult was it for you to formulate a response to deal with the situation created by the new information?

|_________|_________|_________|_________|_________|_________|_________|

not difficult at all	extremely difficult

3. To what extent did you need to modify your COA to accomodate the new information?

|_________|_________|_________|_________|_________|_________|_________|

not at all	to a great extent

4. How much uncertainty is there in the tactical situation <u>now</u>?

|_________|_________|_________|_________|_________|_________|_________|

no uncertainty at all	extremely high uncertainty

5. How confident are you that your response can deal with the situation created by the new information?

|_________|_________|_________|_________|_________|_________|_________|

not confident at all	extremely confident

End of Experiment Questionnaire

Subject ID ____ Date _________

Part I
Background Information

The purpose of this questionnaire is to obtain some information about your military background and experiences. This information will be used to better understand your responses. All information collected will remain confidential and will not be released to third parties. We appreciate your cooperation in completing this form.

Rank________________ Branch: _______

Time in Grade: _______ Time in Service: _______

Last Service School Attended: ____________________________ Year: _______

1. Please indicate all the tactical command or staff positions you have held:

Echelon	Unit Type	Position	Months in Position
__________	______	______________________________________	________
__________	______	______________________________________	________
__________	______	______________________________________	________
__________	______	______________________________________	________
__________	______	______________________________________	________
__________	______	______________________________________	________

2. Please indicate your assignments to units that had operational missions in the Persian Gulf Area.

Position	Unit	Time in Position
____________	____	________
____________	____	________
____________	____	________
____________	____	________
____________	____	________
____________	____	________

3. Please list the training exercises you have participated in (CPX & FTX) that involved the Persian Gulf area:

____

____

____

____

4. To what extent do you consider yourself a student of military history?

|__________|__________|__________|__________|__________|__________|__________|

not at all — to a great extent

5. What aspect of your training or experience did you find most relevant for the situations in this experiment?

Experimenter's Questions

1. Tactical Situation
2. New Information
3. End of Experiment

Experimenter's Questions about the Tactical Situation

1. What was the most critical aspect of the tactical situation?
2. What was your most critical uncertainty, initially?
3. What are some other aspects of the situation you considered while formulating your COA?
4. Is there anything in this situation that reminds you of a previous experience?
5. What information was most influential in reaching your COA?
6. What are the "show stoppers" in this situation? That is, how can your COA be thwarted?
7. How many alternative COAs did you consider?

   Can you briefly explain one alternative COA you considered?

Experimenter's Questions about the New Information

1. What was the most critical aspect of the new information that led you to this response?

2. Were there any other aspects of the situation that you considered while formulating your response to this new information?

End of Experiment Questions

1. Which of the three tactical situations did you consider to be the most complex or difficult? Why?

2. In which tactical situation did you feel you had to deal with the most uncertainty or ambiguity ? Why?

3. Overall, how realistic are the tactical situations posed in this experiment? Which situation was least realistic?

4. In general, were you given enough time to ask questions?

5. On the whole, did you feel you had sufficient time to consider the information and formulate a plan?

6. On average, did you feel you had sufficient time to develop your COA and write out your messages?

7. In general, what were the most important factor influencing your decisions? Why? What other factors played an important role?

8. What information was not useful to you in making your decisions? Why wasn't it helpful?

9. Was there any information that should have been covered in the background (at home) materials that was missing? If yes, please explain.

10. Was there anything about the experiment that was unclear or that we should have explained better?